# The Identification of the Prox Gene Family and the Development of Optical Projection Tomography for Use in Zebrafish

Robert James Bryson-Richardson

PhD

The University of Edinburgh

2003

## Declaration

I declare that:

a) I composed this thesis myself.
b) Unless otherwise acknowledged this work is my own.
c) This work has not been submitted for any other degree or professional qualification

Robert J Bryson-Richardson
Febuary 2003

## Acknowledgements

Firstly I would like to thank my supervisor Pete Currie for the encouragement and tuition he gave me. I am grateful to all the members of the lab; Dave Bassett, Fernando Cortes, Dave Daggett, Lynn Haines, Georgina Hollway, Dave Keenan, John Maule, and Christine Neyt; for their help. I would also like to thank Christine and Dave for their sectioning and sequencing, and Pete, Lynn, Georgina, and Rob Watson for offering their comments on this thesis.

I would like to thank all of the other people at the unit who have taken the time to help me, in particular Martin Taylor and Colin Semple for help with bioinformatics and James Sharpe, Bill Hill, Paul Perry, Richard Baldock, and Margaret Stark for their advice and help with OPT.

I have greatly enjoyed the time I have spent doing my PhD, which is largely due to the friendship, good humour, and company of all those mentioned above, together with my friends and fellow students. Last but not least I would like to thank; my mother, Russ, Miranda, and my family for their limitless supply of encouragement.

## Abstract

Vertebrate homologues of the *Drosophila melanogaster* gene *prospero* have been identified in many species. Whilst the function and regulation of *prospero* has been well studied in *Drosophila* the function and regulation of the homologous vertebrate gene, p*rox*1, is not known.

We describe the identification of the *prox* genes as members of a multigene family in vertebrates through the isolation of new members of the Prox gene family in zebrafish, *Fugu rubripes*, *Tetraodon nigroviridis*, mouse, and human. We examined the phylogeny of this new multigene family and we characterised the expression of these novel genes in zebrafish.

Analysis of the expression of these genes identified the slow muscle as site of expression for *prox*1 that did not overlap with the novel zebrafish Prox genes. Therefore, we studied the function of *prox*1 in the slow muscle using a combination of DNA, and morpholino injections. We demonstrate that *prox*1 in not required for the specification of slow muscle as determined by the expression of markers of terminal differentiation. We also show that the medial lateral migration of the slow muscle is unaffected by the loss of *prox*1. However, ectopic expression of *prox*1 specifically in the fast muscle causes a defect in nuclear patterning. In normal development the fast muscle cells fuse early to form a multinucleate syncytium. The nuclei in this syncytium are normally evenly spaced. Ectopic expression of *prox*1 resulted in the nuclei of the fast cells being positioned at the centre of the syncytium similarly to the situation observed in the mononucleate slow muscle. Furthermore loss of Prox1 results in the disrupted patterning of the slow fibres, demonstrating a role for Prox1 in the patterning of the slow muscle fibres.

An understanding of the 3-dimensional (3D) pattern of gene expression can often lead to a better understanding of gene function. Optical projection tomography

(OPT) is a new method for obtaining 3D data about an object. OPT generates a 3D digital model of a sample and allows it to be virtually sectioned, or rendered to produce a 3D image. OPT was developed for use on mouse embryos and had not been tested with zebrafish. We describe the difficulties of using OPT on samples as small as zebrafish embryos and the development of techniques to overcome these problems and allow its use in zebrafish.

# Table of Contents

# Table of Figures

## Abbreviations

| | |
|---|---|
| 2D | two dimensional |
| 3D | three dimensional |
| aa | amino acid |
| AHD | atypical homeodomain |
| ALD | asymmetric localisation domain |
| *ants* | *antisocial* |
| *ase* | *asense* |
| BLAST | basic local alignment search tool |
| *blow* | *blown fuse* |
| BMP | bone morphogenetic protein |
| Boss | bride of sevenless |
| bp | base pairs |
| cdk | cyclin dependent kinase |
| cDNA | complementary deoxyribonucleic acid |
| CNS | central nervous system |
| *cycA* | *cyclinA* |
| *cycE* | *cyclinE* |
| *dap* | *dacapo* |
| DER | *Drosophila* epidermal growth factor receptor |
| DLST | dihydrolipoamide succinyltransferase |
| DNA | deoxyribonucleic acid |
| *dpn* | *deadpan* |
| *duf* | *dumbfounded* |
| *ehh* | *echidna hedgehog* |
| *en* | *engrailed* |
| EST | expressed sequence tag |
| *eve* | *even-skipped* |
| Fig. | figure |
| *flh* | *floating head* |
| *ftz* | *fushi tarazu* |

| | |
|---|---|
| *gfl* | *gleeful* |
| GMC | ganglion mother cell |
| *hbs* | *hibris* |
| HMM | hidden Markov model |
| *insc* | *inscuteable* |
| IRES | internal ribosome entry site |
| kb | 1000 base pairs |
| *lsc* | *lethal of scute* |
| Lz | Lozenge |
| MAPK | mitogen-activated protein kinase |
| *mbc* | *myoblast city* |
| *mef-2* | *myocyte enhancing factor 2* |
| *mira* | *miranda* |
| *mlc2f* | *myosin light chain 2 fast* |
| MP | muscle pioneer |
| mSos | mammalian son-of-sevenless |
| MyHC | myosin heavy chain |
| MyoD | Myogenic differentiation antigen 1 |
| NB | neuroblast |
| NES | nuclear export sequence |
| NLS | nuclear localisation sequence |
| *ntl* | *no tail* |
| OPT | optical projection tomography |
| PCR | polymerase chain reaction |
| PD | *prospero* domain |
| PFA | paraformaldehyde |
| Pfu | plaque forming unit |
| PI | propidium iodide |
| PNS | peripheral nervous system |
| Pnt | Pointed |
| Prox | *prospero*-related homeobox |
| RNA | ribonucleic acid |

| | |
|---|---|
| rpm | revolutions per minute |
| *rst* | *roughest* |
| *sev* | *sevenless* |
| *shh* | *sonic hedgehog* |
| *sina* | *sevenless in absentia* |
| SMC | sensory mother cell |
| *smu* | *slow-muscle-omitted* |
| sMyHC | slow myosin heavy chain |
| *sns* | *sticks and stones* |
| *stau* | *staufen* |
| *stg* | *string* |
| TGF | transforming growth factor |
| Ttk88 | Tramtrack |
| VTK | the visualization toolkit |

# Chapter 1: Introduction

## 1.1 Muscle Formation in the Zebrafish

Zebrafish have many advantages for the study of muscle development. The optically clear embryo of the zebrafish is ideal for microscopic examination. Large mutational screens have also been carried out (Driever et al., 1996; Haffter et al., 1996), identifying many mutants affecting the formation of muscle. In particular the u-class mutants, so named because of their characteristic blocky u-shaped somites compared to the chevrons of wildtype embryos (van Eeden et al., 1996). The characterisation of mutants from these screens and the identification of the genes involved, together with the use of other strategies involving the analysis of cloned genes *in vivo* have led to an increased understanding of muscle formation.

In the zebrafish the skeletal muscle of the trunk is formed from the myotome, which constitutes the bulk of the somites. The somites are derived from the presomitic mesoderm and form in an anterior to posterior progression, one somite forming every 20-30 minutes up to approximately 30 somites (for a review of somitogenesis see Stickney et al., 2000). All vertebrate skeletal muscles are composed of two specific types of fibres, slow and fast twitch. Slow, or red, muscle has a long contraction period with a small force and is used for aerobic activity. Fast, or white, muscle has a rapid, high force contraction used in sudden movement such as escape response. The two fibre types can be characterised by the expression of slow or fast myosin heavy chains (MyHC). In higher vertebrates the two fibre types are interspersed with each other throughout the myotome. In zebrafish, and other teleost fishes, the embryonic origins of slow and fast muscle fibres are discrete (Blagden et al., 1997), and the two fibre types are spatially separated (Waterman, 1969; van Raamsdonk et al., 1978).

The commitment of cells to myogenic fates is characterised by the onset of *MyoD* expression which commences in the presomitic mesoderm, and is initially confined to a triangular pattern in each forming somite either side of the notochord (Weinberg et al., 1996). Later in development, the cells adjacent to the notochord, the adaxial cells, will adopt a distinct cuboidal appearance (Fig. 1.1a). These *MyoD* expressing adaxial cells (Fig.1.1b) will eventually go on to form the slow muscle of the zebrafish, including the first differentiating muscle cells, the muscle pioneer (MP) cells (Fig.1.1c; Felsenfeld et al., 1991).

The MP cells elongate at the midline, near the future horizontal myoseptum. Of the approximately 20 adaxial cells of each somite between three and six will go on to form MP cells. These cells are labelled by the expression of *engrailed* 1 and 2 (Ekker et al., 1992). *engrailed* (*en*) is a homeodomain containing transcription factor identified in *Drosophila*, where it controls segment polarity and is also involved in neurogenesis. *en*-1 and *en*-2 are visible two to four somites anterior to the newly formed somite. *en*-2 is transiently expressed, *en*-1 persisting late into development (32hr+; Ekker et al., 1992). Differentiation of the muscle pioneers coincides with the transition of the zebrafish somites from a blocky to chevron appearance.

Work carried out on the *no tail* (*ntl*) and *floating head* (*flh*) mutants, which lack the notochord, has shown that the notochord is necessary for the formation of the MP cells (Halpern et al., 1993; Talbot et al., 1995). The developing somites of *ntl* and *flh* mutants lack muscle pioneer cells, fail to adopt a chevron shape, and do not form myosepta. Transplanted notochord cells were able to rescue the phenotype but only in the cells immediately adjacent to the wildtype transplant (Halpern et al., 1993). This work has shown that signalling from the notochord is necessary for the correct specification of the MP cells. It is not known how the MP cells are required in the chevron patterning of the somites and the formation of the horizontal myospeta, but these defects have been observed in other mutants lacking MP cells where the notochord is present, such as *u-boot* (Roy et al., 2001).

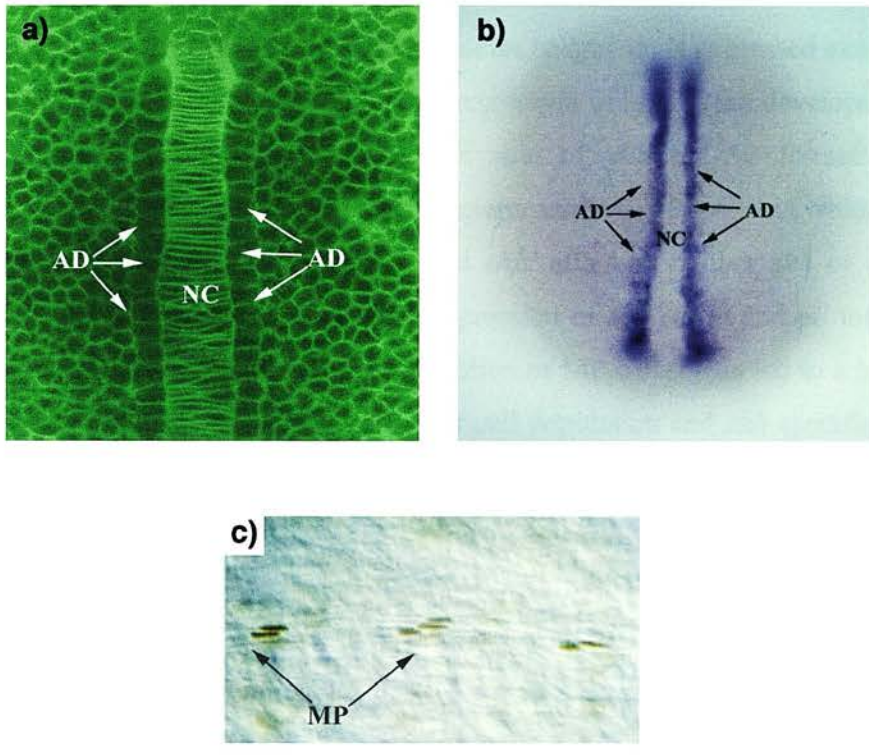



**Fig. 1.1 Adaxial cell morphology and expression of *MyoD***

a) Dorsal view of adaxial cells (AD) flanking the notochord (NC) in a bodipy ceramide stained embryo at the segmentation stage. The cuboidal appearance of the adaxial cells is clearly in contrast to the rest of the mesoderm.
b) Adaxial cell expression of *MyoD* can be detected by in situ hybridisation in a pre-segmentation stage embryo.
c) Lateral view of a 26-somite stage embryo labelled using an antibody against Engrailed protein (brown) which is expressed within the nuclei of muscle pioneer cells (MP).
This diagram is taken from Currie and Ingham 2001.

Among the proteins expressed in the notochord are two members of the hedgehog secreted glycoprotein gene family, *sonic hedgehog* (*shh;* Krauss et al., 1993) and *echidna hedgehog* (*ehh;* Currie and Ingham, 1996). *shh* is expressed in the *ntl* and *flh* mutants, but *ehh* is absent. *sonic you,* a *shh* mutant, has absent or reduced MP cells, and also shows a reduction it the amount of slow muscle (Schauerte et al., 1998).

*ehh*, which is most closely related to *indian hedgehog*, is expressed exclusively in the notochord. This is in contrast to the expression of *shh* in the developing notochord, floor plate, limb bud and eye. The role of *ehh* in MP formation is further demonstrated by the induction of supernumerary MP cells in wildtype embryos following over expression of *shh* and *ehh,* although neither *shh* or *ehh* alone are sufficient to induce MP cells. The expression of *ehh* in *flh* and *ntl* mutant embryos was sufficient to cause a partial restoration of MP cells. This led to a hypothesis that ectopic *shh* expands the MP precursor cell population and *ehh* specifies the MP cell fate (Currie and Ingham, 1996).

The adaxial cells that do not form MP cells will form the remainder of the slow muscle population. The non-pioneer slow cells undergo a lateral migration from their position adjacent to the notochord to form a superficial layer of subcutaneous muscle (Fig. 1.2). This migration occurs in an anterior to posterior fashion similar to somitogenesis and commences at around the 20-somite stage in anterior somites (Devoto et al., 1996; Blagden et al., 1997).

Whilst it appears that the role of *ehh* is limited to specification of MP cells, *shh* is necessary for the formation of all slow muscle. In the *youtoo* mutant, a mutation in the Gli2 gene, which encodes a zinc finger transcription factor that is the effector of hedgehog signalling, there is almost no slow muscle and a complete loss of MP cells (Karlstrom et al., 1999). Ectopic expression of *shh* has also been shown to cause an expansion of slow muscle at the expense of fast both *in vivo* (Blagden et al., 1997) and in culture (Norris et al., 2000). *shh* expression has also been shown to be able to induce *MyoD* in chicks (Borycki et al., 1998). The *slow-muscle-omitted* mutant *smu* demonstrates a complete lack of MP cells and a 99% reduction in non-pioneer slow

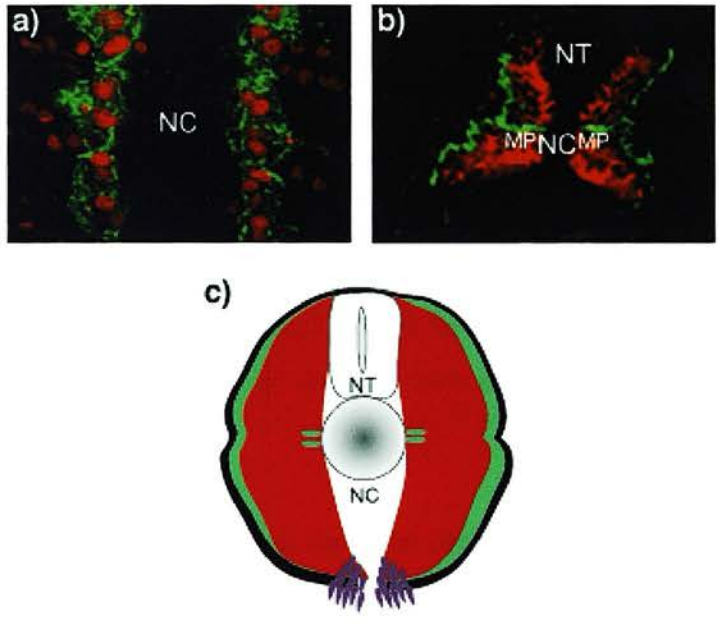



**Fig. 1.2 Migration of the non-pioneer slow muscle**

a) A confocal microscope view of a 15-somite stage embryo. The adxial cells are labelled with a slow muscle specific antibody (green), and myogenic transcription factor positive nuclei (Red). Dorsal view, anterior to the top (NC, notochord).
b) A confocal microscope cross-section. Double antibody staining reveals that slow muscle (Green) migrates to the lateral extent of the myotome (except for the muscle pioneer cells, MP) and fast muscle (Red) differentiates behind this wave of migration to form the remainder of the muscle cells of the myotome. (NT, neural tube).
c) Schematic cross-section of a zebrafish embryo revealing that the mature myotome, with the separate populations of fast muscle (red) and slow muscle (green) constitutes the bulk of the somitic derivatives while the sclerotome (purple) is relatively small in comparison to amniotes.
Diagram taken from Currie and Ingham 2001.

fibres. *smu* is a mutant in the seven transmembrane G-protein coupled receptor *smoothened* (Varga et al., 2001; Chen et al., 2001), an essential component in signalling from all of the hedgehog proteins (van den Heuvel and Ingham, 1996).

Other research has suggested that *shh* and *tiggy-winkle hedgehog* are capable of inducing MP cells when overexpressed alone, not just a combination of *ehh* and *shh* (Du et al., 1997). Du and colleagues also proposed that signalling from members of the TGF-β family inhibits MP cell formation. Ectopic expression of chick Dorsalin-1, a BMP4-related protein, in the notochord prevented the formation of MP cells by inhibiting hedgehog signalling downstream of PKA activity. The specification of non-pioneer slow muscle was unaffected. Therefore, they propose it is a combination of signals from Hedgehog and TGF-β family members that determine slow muscle cell type. It remains to be seen if there is a BMP4 like signal in the zebrafish myotome.

Whilst Hedgehog signalling has been clearly demonstrated to be involved in the formation of embryonic slow muscle it has also been demonstrated that formation of new slow muscle fibres at the larval stage occurs independently of Hedgehog signalling (Barresi et al., 2001).

The remaining cells of the myotome fuse and differentiates as fast muscle in a wave behind the advancing front of migrating slow muscle. The differentiation of fast muscle however does not depend on the presence of slow muscle as fast muscle differentiates at the same time point in embryos completely lacking in slow muscle (Blagden et al., 1997).

Whilst the morphogenesis of the slow muscle has been described (Waterman, 1969; Raamsdonk et al., 1974; Devoto et al., 1996), the molecular signals downstream of hedgehog signalling required for the specification of slow muscle are not known. Glasgow and Tomarev (1998) isolated a transcription factor, *prox*1, which was specifically expressed in the slow muscle and its precursors. *prox1* is the vertebrate homologue of the *Drosophila prospero* gene. The expression of *prox*1 in the

zebrafish is initially restricted to the lens primordium and in the MP cells. Later in development (28h) expression is more widespread; including the hindbrain, otic vesicle, lateral line primordium, as well as the MP cells and the non-pioneer slow muscle cells. This was the first transcription factor identified that was restricted in the myotome to the slow muscle and therefore we wished to determine the role of Prox1 in slow muscle development.

## 1.2 *prospero*

The product of the *Drosophila melanogaster prospero* gene, Pros, is a homeodomain containing transcription factor and the founding member of the *prospero*-related homeobox (or Prox) protein family.

### 1.2.1 *prospero* Isolation and Expression

*prospero* mutations (*pros*) were originally identified in two *Drosophila* screens; one for altered ganglion mother cell (GMC) and neuron fates, and an enhancer trap screen for neuroblast genes. Doe *et al.* identified seven alleles of *pros*, all of which were embryonic, or embryonic/larval, lethal. The enhancer trap was used to map *pros* to region 86E on chromosome 3R (Doe et al., 1991; Vaessin et al., 1991).

*pros* is uniformly expressed at the blastoderm stage. By the gastrula stage, expression is restricted to seven stripes, with higher expression in the precursors of the brain neuroblasts, the lateral cephalic cells (Doe et al., 1991). In the central nervous system (CNS) *pros* is expressed in all but two of the neuroblasts in each hemisegment and many of the GMCs. In the peripheral nervous system (PNS) expression is initially observed in a single sensory mother cell (SMC; Chu-LaGraff et al., 1991). Each SMC will divide to form the neuron and non-neuronal cells of the sensory organ along a set lineage (Bodmer et al., 1989). After 12 hours of development expression

is restricted to a single non-neuronal cell in the mature sensory organ (Chu-LaGraff et al., 1991).

### 1.2.2 The *pros* Phenotype

*pros* was identified through the alteration of GMC fate in the mutant. No change was observed in the number of neuroblasts but a subset developed along a different cell lineage to the wild type (Doe et al., 1991). The expression of the segmentation genes *fushi tarazu* (*ftz*; Doe et al., 1988a) and *even-skipped* (*eve*; Doe et al., 1988b) is lost in some of the GMC whilst that of the homeobox gene *engrailed* (*en*) is expanded, resulting in a change of neuron identity (Doe et al., 1991). The expression of *ftz*, *eve*, and *en* was unaffected at the segmentation stage and no segmentation defects were observed. The loss of *ftz* and *eve* in GMCs has been shown to produce axon pathfinding defects in the daughter cells (Doe et al., 1988a; Doe et al., 1988b) and defects in axon pathfinding was observed in *pros* mutant embryos (Doe et al., 1991). Vaessin *et al.* 1991 examined the expression of the neuronal precursor genes *deadpan* (*dpn*; Bier et al., 1992) and *asense* (*ase*; Gonzalez et al., 1989) in *pros* mutants and found, contrary to the wild type, expression was not restricted to a subset of cells in the CNS. In addition, expression in both the CNS and PNS persisted until later in development compared to the wildtype.

Muscle defects have been described in *pros* embryos although there is no muscle expression of *pros*. These defects are due to a reduction of synaptic receptors coupled with a failure of the receptors to localise at the neuromuscular junction. This is a result of delay or failure of innervation, and is a secondary phenotype to the axon defect (Broadie and Bate, 1993).

Increased mitotic activity in *pros* embryos was detected by an increase in Histone 1A and phosphorylated Histone H3, and expression of cell-cycle genes *cyclin A* (*cycA*), *cyclin E* (*cycE*), and *string* (*stg;* Li and Vaessin, 2000). The increased mitotic activity was potentially compensated for by cell death, as suggested by the increased

expression of *reaper* (Li and Vaessin, 2000), a marker of apoptosis. Similarly, ectopic Pros expression was found to reduce transcription of *stg*, *retinoblastoma-family protein*, *E2F transcription factor*, *cycA*, and *cycE*. It is not known if this effect is direct, or due to a downstream effector of Pros. Pros is known to inhibit *dpn* (Vaessin et al., 1991), which is itself an inhibitor of the cyclin dependent kinase (cdk) inhibitor gene *dacapo I* (*dap*) necessary for cells to enter their final cell division (Wallace et al., 2000).

The examination of the *pros* mutants demonstrates a role for determining neuron identity and regulating proliferation in a subset of the neuronal cell lineage.

### 1.2.3 *prospero* is Asymmetrically Localised

Pros is asymmetrically localised following neuroblast division (Hirata et al., 1995). Whilst it is transcribed in both the neuroblast and the GMC, the translation product is only localised to the nucleus in the GMC (Spana and Doe, 1995). Other asymmetrically localising genes have been identified, but in these cases asymmetric localisation is dependent on RNA sequence. Pros was the first protein identified with an asymmetric localisation domain (Hirata et al., 1995). Although Pros has a protein localisation domain it has also been shown that the RNA is asymmetrically localised (Li et al., 1997). Both the protein and RNA form a crescent at the basal cell membrane of the neuroblast and remain localised at the membrane through metaphase and anaphase, before segregating to the GMC at the end of telophase (Fig. 1.3; Spana and Doe, 1995).

Three genes have been identified which are important in the localisation of the Pros protein and RNA. *inscuteable* (*insc*) is required for the localisation of both the protein and RNA at the basal cortex. It also has a role in the orientation of the mitotic spindle (Kraut et al., 1996). *staufen* (*stau)* was isolated from a screen for proteins that bind Insc. *staufen* had previously been shown to be involved in the localisation of *bicoid* and *oskar* in the oocyte (St.Johnston et al., 1989; Ferrandon et al., 1994).

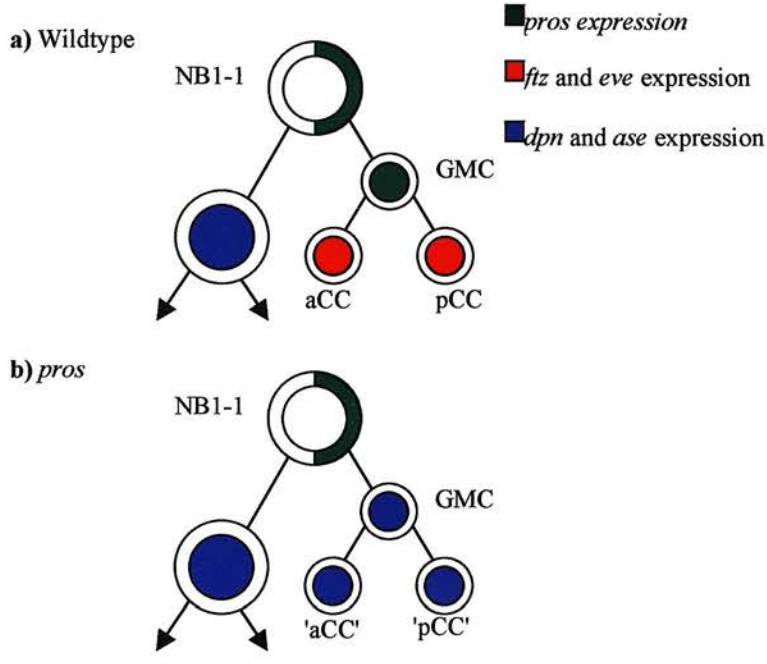



**Fig. 1.3 Asymmetric *prospero* localisation**

a) *prospero* is expressed in neuroblast 1-1 (NB 1-1) and is asymmetrically localised in the daughter cells. In the GMC Pros is translocated from the cortex to the nucleus. The GMC and its progeny, the aCC and pCC neurons, express homeobox genes *ftz* and *eve*. Neuroblasts express the neural precursor gene *deadpan* (*dpn*) and *asense* (*ase*).
b) In the *pros* mutant the neural precursor genes are not repressed, and the GMC genes *eve* and *ftz* are not expressed.
Adapted from Doe *et al.* (1991) and Spana and Doe (1995).

*staufen* was shown to act downstream of *insc* but was not necessary for mitotic spindle formation or Pros localisation. In the *stau* mutant only mRNA localisation was affected. However there is no shift in GMC fate (Li et al., 1997; Broadus et al., 1998), showing that the correct localisation of the protein is sufficient for wild type fate.

The third gene identified, *miranda* (*mira*), is necessary for both protein and mRNA localisation. Mira interacts with Pros via its asymmetric localisation domain, and was identified through this interaction. *miranda* is expressed in asymmetrically dividing cells coincident with *prospero* expression, and GMC cell fate is altered in *mira* mutants. *miranda* was also found to act downstream of *insc* and its localisation is not affected in the *pros* mutant (Shen et al., 1997). Once Pros is segregated into the GMC *mira* is no longer detectable and Pros moves to the nucleus. Overexpression and mutant studies have shown that Pros is nuclear localised even if Mira is present, suggesting that degradation is not necessary for Pros release, instead a phosphorylation domain in the C-terminal of Mira may be important for Pros release (Ikeshima-Kataoka et al., 1997). The role of Mira in RNA localisation is due to its binding with Stau (Shen et al., 1998). Shen and colleagues also demonstrated that the asymmetric localisation domain of Mira interacts with Insc. This same mechanism was also shown to function in the asymmetric localisation of Pros in epithelial cells (Matsuzaki et al., 1998). Recent work has also demonstrated that nuclear localisation is not solely dependent on Mira. Pros contains a nuclear export sequence (NES) the function of which is masked by the Prospero Domain (PD). Removal of the PD results in cytoplasmic localisation due to the functional NES (Demidenko et al., 2001). How the PD controls the NES is not clear although the authors suggest that phosphorylation of the PD may regulate its function and it has been demonstrated that cytoplasmic Pros is more highly phosphorylated than nuclear Pros (Srinivasan et al., 1998).

### 1.2.4 Genomic Structure of *prospero*

*prospero* maps to chromosome 3 and is encoded by 5 exons spanning 19kb. There are two alternative transcripts of *prospero*, a long form, prosL, and a short from, prosS. The only difference between the two forms is 87bp of coding sequence, which results in a 29 amino acid insertion in front of the homeobox region (Chu-LaGraff et al., 1991). The alternative transcripts result from two nested splice sites (Fig. 1.4). The long cDNA, prosL, results from splicing at the conventional U2-type intron. The short from, prosS, results from splicing of a U12-type intron flanking the U2 intron such that it removes 59 nucleotides 5' and 28 nucleotides 3' of the U2 splice site from prosL (Otake et al., 2002). U12 introns are rare introns that are spliced by the U12 spliceosome and have an AT-AC splice site (Jackson, 1991; Hall and Padgett, 1994). Otake *et al.* observed that the ratio of prosS to prosL in wildtype is 3.6:1 and concluded that the U12 intron is favoured compared to the U2. However nested splice sites form 'twintron' structures (Copertino and Hallick, 1991) in which the internal site is spliced first followed by the external splice site, thus in all cases the U2 intron is spliced. The ratio of ProsS to ProsL instead demonstrates that the U12 intron is spliced in approximately 80% of transcripts. It has recently been suggested that U12 introns may be a form of post-transcriptional regulation, as they are spliced much more slowly than U2 introns (Patel et al., 2002). Splicing of the U12 intron may be another layer of regulation in the control of *prospero* expression. Analysis of a U12 type spliceosome mutant suggested that prosL alone is sufficient for early development (Otake et al., 2002).

### 1.2.5 Pros Protein Structure

Pros has many functional domains that have been defined by sequence similarity and experimental analysis. Prospero contains an atypical homeodomain at amino acids (aa) 1241-1303, although atypical, threading analysis predicts that the homeodomain is still functional (Banerjee-Basu et al., 1999). In the short form of Pros, ProsS, aa

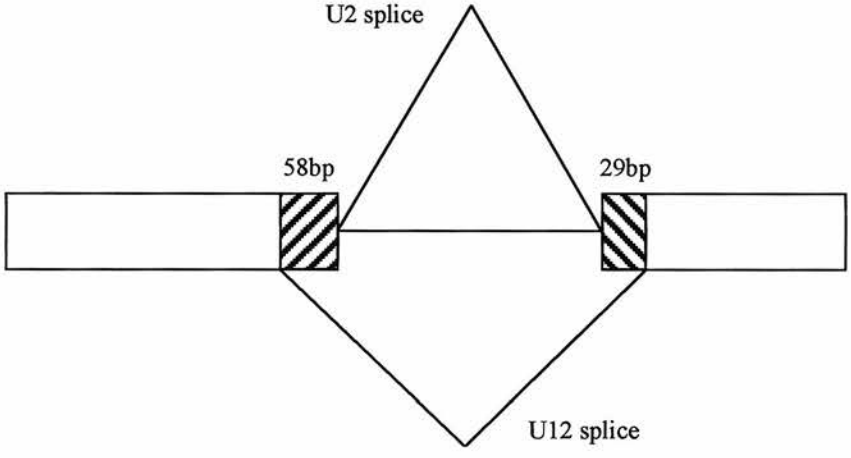



**Fig. 1.4 Splicing of *prospero***

The twintron of *prospero* contains an inner U2 dependent splice site and an outer U12 dependent splice site. Adapted from (Otake et al., 2002).

1216-1244 are removed. This results in the loss of the first four amino acids of the homeodomain. The *prospero* domain which, together with the atypical homeodomain, defines the members of the Prox gene family was identified by sequence conservation between Pros and the *C.elegans* homologue (Burglin, 1994), and is at aa 1304-1403 in *Pros*. The C-terminal 30 amino acids of the *prospero* domain was later shown to be sufficient to mask the function of a nuclear export sequence which was mapped to amino acids 1252-1265 (Demidenko et al., 2001). Masking of the NES allows nuclear localisation, the NLS mapping to aa 991-998. The domain required for asymmetric localisation was mapped to amino acids 871-902 (Hirata et al., 1995). The structure of Pros is shown in Fig.1.5.

### 1.2.6 Regulation of *prospero*

The regulation of *prospero* has been investigated in the R7 equivalence group of the ommatidia in the developing eye. The development of the ommatidia is reviewed in Tomlinson (1988). In each ommatidium the eight photoreceptor cells form in a set order; R8, R2 and R5, R3 and R4, R1 and R6, and finally R7. After the R7 photoreceptor has been added four lens-secreting cone cells are also recruited to the cluster. These four cone cells are also capable of differentiating into an R7 photoreceptor, and therefore the R7 precursor and the cone cell precursors are referred to as the R7 equivalence group. Analysis of the *sevenless* (*sev*) mutant demonstrated a cell autonomous defect resulting in the loss of R7 (Harris et al., 1976; Campos-Ortega et al., 1979). The *sev* gene was later identified to encode a transmembrane receptor tyrosine kinase (Banerjee et al., 1987; Hafen et al., 1987)). The specification of R7 requires signalling from Bride of sevenless (Boss; Reinke and Zipursky, 1988) on the surface of the adjacent R8 photoreceptor to Sev on the surface of the R7. This results in the activation of the Sev/Ras1/Raf/MAPK signalling pathway (Blake et al., 1992; Tsuda et al., 1993), activation of which results in ectopic R7 photoreceptors (Fortini et al., 1992).

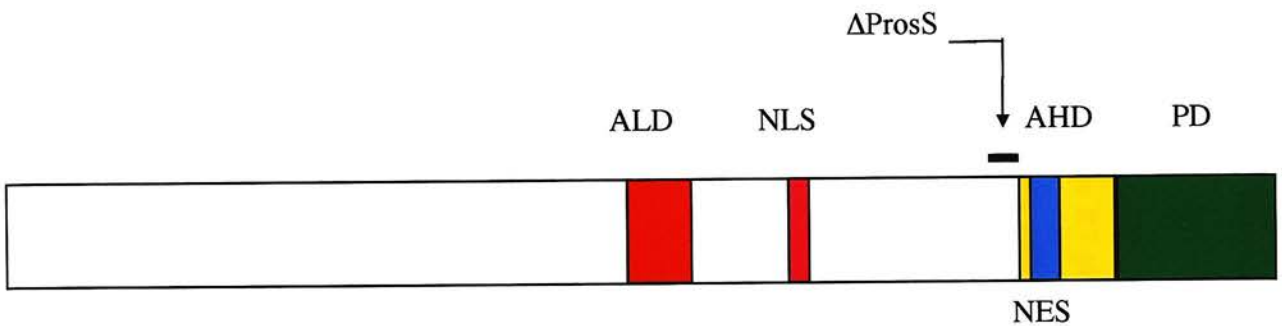



**Fig. 1.5 Pros protein structure**

Five domains have been identified in the Pros protein, an asymmetric localisation domain (ALD, red), nuclear localisation signal (NLS, purple), nuclear export sequence (NES, blue), atypical homeodomain (AHD, yellow), and the *prospero* domain (PD, green). The region deleted in the ProsS variant is also shown.

*pros* is expressed uniformly in the R7 equivalence group before being upregulated specifically in the R7 photoreceptor (Kauffmann et al., 1996). In *sevenless in absentia* (*sina*) mutants the R7 photoreceptor precursor develops as a cone cell (Carthew and Rubin, 1990). *pros* was identified as a dominant enhancer of a weak allele of *sina* (*sina*$^{4}$), in which only 20% of R7 were missing. In the *pros/sina*$^{4}$ double mutant half of the ommatidia lacked an R7 (Kauffmann et al., 1996). Examination of the R7 in *pros* mutants demonstrated that they had disordered axon projections and in 23% of cases the cell had an altered morphology. However in *sev* mutants *pros* is initially expressed at wild type levels in the R7 equivalence group, but is not upregulated in the R7 photoreceptor precursor, and the R7 photoreceptor precursor develops as a cone cell. Conversely elevated *pros* expression in the R7 equivalence group is sufficient to cause cone cells to be transformed to a R7 cell fate (Kauffmann et al., 1996). This demonstrates that Sev signalling is required for elevated levels of *pros* in the R7 but not for the lower levels of expression throughout the R7 equivalence group.

Xu *et al.* (2000) demonstrated that *pros* is also regulated by *Drosophila* epidermal growth factor receptor (DER; Livneh et al., 1985). Like Sev, DER functions through a Ras1/Raf/MAPK transduction pathway, and the two proteins have been demonstrated to be functionally equivalent (Freeman, 1996). Pointed (Pnt) and Yan are targets of the activated MAPK, phosphorylation by MAPK activates Pnt and decreases the negative regulatory effect of its competitor Yan (O'Neill et al., 1994). Analysis of the *prospero* promoter demonstrated that Yan and Pnt directly bound an upstream enhancer (Xu et al., 2000). However Xu and colleagues also demonstrated that Yan and Pnt are not sufficient for *prospero* expression, requiring the presence of the Ras1 independent transcription factor Lozenge (lz; Daga et al., 1996; Flores et al., 1998), expression of Lz and DER being sufficient for *pros* expression. The promoter region containing the Yan and Pnt binding sites drove low level expression throughout the R7 equivalence group, but not the elevated expression in the R7 precursor. Further analysis of the promoter region demonstrated that it was not responsive to Sev. The addition of binding sites for Tramtrack (Ttk88), a repressor of the R7 cell fate (Lai et al., 1996), resulted in upregulation in response to Sev

signalling suggesting that Sev regulates *pros* transcription through relief of Ttk88 inhibition.

### 1.2.7 Targets of *prospero*

In the *pros* mutant, the neural precursor genes *dpn* and *ase* are expressed in the GMC and its daughter cells, and *ftz* and *eve* are lost. The function of *ftz* and *eve* in these cells is unknown. It has been suggested however, that homeoproteins have a very wide range of targets regulating, either directly or indirectly, 87% of genes in late embryogenesis (Liang and Biggin, 1998). At the cellular blastoderm stage it was also suggested that *ftz* and *eve* may regulate up to 50% of transcripts expressed and, given the rapid response to *ftz* and *eve* expression, many or all of these may be directly regulated.

## 1.3 The Vertebrate Prox Genes

*prox1* is the vertebrate homologue of the *Drosophila prospero* gene. *prospero* homologues have been identified in other species through the conservation of the homeodomain and *prospero* domain (PD), including mouse (Oliver et al., 1993), *C.elegans* (Burglin, 1994), chick (Tomarev et al., 1996), and human (Zinovieva et al., 1996). There is little conservation of the N-terminal end of the protein between *Drosophila* and *C.elegans* Prox proteins compared to the zebrafish, human, chick, and mouse Prospero homologues which are closely related.

In the chick, where *prox1* is expressed in the lens, retina, midgut, and liver and pancreatic diverticulum there is evidence for alternative splicing of *prox1* as three RNAs were isolated of 2kb, 3.5kb, and 8kb. Southern hybridisation was only able to detect the presence of one *prox1* gene in chick (Tomarev et al., 1996). The only experimental data for the function of *prox1* in vertebrates comes from the mouse.

### 1.3.1 Prox1 Knockout Mice

The study of *prox1* in the mouse has shown expression in the lens, pancreas, neurons of the subventricular region, liver, heart, and transiently in skeletal muscle (Oliver et al., 1993) Studies of the mouse knockout have illustrated many roles for *prox1* in development. The *prox1* null embryos die of multiple defects at mid-gestation, and have been studied with regard to the phenotype in the lens, liver, and lymphatic system.

In a wildtype embryo, E-cadherin, a marker of epithelial cell proliferation, is expressed throughout the lens vesicle up to E12.5, the time of fibre-cell elongation, then becomes restricted to the anterior proliferative epithelium. At the same time, non-proliferating posterior epithelium terminally differentiates into lens fibres. In the absence of *prox1* the cell cycle inhibitors $p27^{KIP1}$ and $p57^{KIP2}$ are down regulated. $prox1^{-/-}$ mice show E-cadherin expression is maintained throughout the lens after E12.5, and lens fibres fail to terminally differentiate and elongate (Wigle et al., 1999).

In the liver *prox1* expression marks the hepatic primordium and dorsal pancreatic bud at E9.5, spreading to the hepatic bud, gall bladder, and dorsal and ventral pancreatic primordia at E10.5. In a wildtype embryo, hepatocytes delaminate from the hepatic epithelium in an area where the basal membrane is lost and migrate laterally. High levels of E-cadherin are found in hepatocytes delaminating from the hepatic epithelium, a lower concentration being expressed in the migrating hepatocytes. Within the line of migrating hepatocytes there are cells which continue to proliferate. In the *prox1* null embryos, the membrane remains continuous, proliferation is reduced, and there is a failure of hepatocyte migration. High levels of E-cadherin are observed in all of the hepatocytes, not just those delaminating from the hepatic epithelium (Sosa-Pineda et al., 2000).

$prox1^{-/-}$ mice also have a defect in the development of the lymphatic system. *prox1* is expressed in a subpopulation of endothelial cells in the embryonic veins. Co-staining

with a lymphatic marker showed that *prox1* is expressed in the endothelial cells that will form the lymph sacs following budding from the veins. The number of cells budding from the veins was greatly reduced in homozygotes compared to wildtype littermates, the lymphatic capillaries being completely absent in the homozygous *prox1*$^{-/-}$ mouse. There is also no polarity of budding in the *prox1* mutant cells. The results indicate *prox1* expression is required to maintain budding and sprouting of the endothelial cells rather than initiation (Wigle and Oliver, 1999).

### 1.3.2 Asymmetric Localisation of Prox1

Until recently, it was thought that the vertebrate Pros homologue, Prox1, did not alter in subcellular localisation. It has now been shown that, in the lens at least, there is shuttling of Prox1 between the nucleus and cytoplasm (Duncan et al., 2002). Duncan and colleagues (2002) examined the expression of Prox1 in the human, rat, and mouse lens, and observed that whilst predominantly cytoplasmic in the lens placode and epithelium, during fibre cell differentiation Prox1 relocates to the nucleus. Localisation within the cytoplasm is not specific as in the case of Pros. This is expected as the asymmetric localisation domain required for cortical localisation of Pros does not appear to be present in Prox1 (Tomarev et al., 1996; Zinovieva et al., 1996). However as suggested by Duncan *et al.* the homeodomain and PD is well conserved and it is likely that the regulation of a nuclear export sequence (NES) by the PD, as identified by Demidenko et al. (2001), in tandem with import due to the nuclear localisation sequence (NLS) regulates cellular localisation.

### 1.3.3 Targets of Prox1

In the eye Prox1 has been demonstrated to be necessary for γ-crystallin gene expression (Lengler et al., 2001). The γ-crystallins are lens specific structural proteins required for the optical clarity of the lens, mutations in which have been

implicated in cataracts (Heon et al., 1999). Lengler and colleagues demonstrated that Prox1 upregulated the murine *Crygf* expression 10-fold in cell culture. Expression of *Crygf* was inhibited by Six3, which acts through a different binding site, and required Sox1, a known activator of the γ-crystallin genes (Nishiguchi et al., 1998), expression in addition to Prox1. Putative *prospero* binding sites were identified in the promoter region but functional analysis narrowed the Prox1 responsive element to 24 base pairs that did not include the putative *prospero* binding site (Lengler et al., 2001).

## 1.4 Prox1 Function in Slow Muscle

The expression of *prox*1 in the slow muscle led us to suggest four hypotheses for its function. Firstly *prox1* may act downstream of hedgehog signalling to specify a slow muscle fate. Pros has been demonstrated to be involved in fate determination in GMC cells in *Drosophila* and Prox1 may be involved in specifying the slow fate in zebrafish. Additionally data from *Drosophila* and the mouse knockout suggests Prox1 may be required to prevent proliferation, a necessary step to allow terminal differentiation to occur.

*prox1* is expressed throughout the slow muscle prior to migration, during migration, and down-regulated shortly after the completion of migration. Prox1 may play a role in the migration of the adaxial cells to form a superficial layer. Defects in budding have been described in the liver and lymphatic system of knockout mice (Sosa-Pineda et al., 2000; Wigle and Oliver, 1999), and in axon pathfinding in *Drosophila*. This may suggest a role in the co-ordination of migration, if not the promotion of migration itself.

The fusion of fast myoblasts occurs behind the advancing wave of slow migration. Prox1 may function to prevent the fusion of myoblasts in the slow muscle, as occurs in the fast muscle, and is responsible for mononucleate nature of slow muscle.

Finally as had been demonstrated in the eye, Prox1 may be required for the regulation of structural proteins.

### 1.4.1 Muscle Migration

No analogous process has been described in higher vertebrates to the migration of medial muscle to the body wall in zebrafish. Migration of muscle from the somites to the limbs has been described in chick and mice (Williams and Ordahl, 1994) and migration from the somites to the fin musculature in zebrafish (Neyt et al., 2000). In these cases the migration occurs prior to terminal differentiation in contrast to the migration of differentiated slow muscle from the midline to the body wall, as has been described in zebrafish (Devoto et al., 1996) and trout (Rescan et al., 2001). Given the budding and migration defects in the mouse knockout, and axon pathfinding in *Drosophila*, Prox1 may be required for the medial lateral migration of the non-pioneer slow muscle. Whilst Prox1 is also expressed in the MP cells, it is possible that *en*, which has been well characterised as a transcriptional repressor (Jaynes and O'Farrell, 1991), prevents migration in the MP cells.

### 1.4.2 Muscle Fusion

The genetics of myoblast fusion in vertebrates is not well described. It is known though, that in zebrafish, the fast muscle cells fuse to from a multinucleate syncytium behind the advancing wave of migrating non-pioneer slow muscle, whilst the slow muscle remains mononucleate. Prox1 is expressed in this migrating slow muscle and it is possible that Prox1 may have a role in preventing the fusion of slow muscle at a time when the rest of the somitic muscle is fusing.

### 1.4.3 Structural Integrity of the Slow Muscle

A role for Prox1 in the structure and patterning of the slow muscle is supported by its regulation of γ-crystallins in the lens (Lengler et al., 2001). The γ-crystallins are structural proteins required for the opacity of the lens. The γ-crystallins are part of a larger family of crystallins, including the α-crystallins and β-crystallins. In addition to a structural role the α-crystallins, αA-crystallin and αB-crystallin, also act as small heat shock proteins. Recently is has been described that zebrafish αA-crystallin is detectable at low levels in the liver and spleen, in addition to the very high levels in the lens (Runkle et al., 2002). Furthermore in the mouse, expression of αB-crystallin has been detected in the skeletal muscle, heart, lungs, kidney, brain and spleen, in addition to strong lens expression. Whilst Prox has only been demonstrated to regulate γ-crystallin, it is conceivable that Prox1 may also regulate other members of the crystallin family. The expression of crystallins is also much higher in the lens than in any other tissue, as is also observed with Prox gene expression. Furthermore, in the skeletal muscle of the mouse and rat, expression is much higher in the slow muscle than in the fast (Atomi et al., 1991; Benjamin et al., 1997; Atomi et al., 2000). Atomi and colleagues suggest that because of the higher metabolic rate and protein turnover in slow fibres, αB-crystallin is required as a chaperone to protect proteins in the slow muscle, acting as a myofibril-stabilising protein. A dystrophic phenotype has been reported in mice which has the genomic region encoding αB-crystallin removed, but unfortunately the myotonic dystrophy protein kinase binding protein (MKBP), an ancient duplication of αB-crystallin (Iwaki et al., 1997), was subsequently discovered to be in the same region (Brady et al., 2001). It is therefore not possible to distinguish if loss of MKBP or loss of a, structural or chaperone, function of αB-crystallin, leads to the dystrophic phenotype.

Prox1 may be required for the structural integrity of the slow muscle through regulation of the crystallins. In particular Prox1 may regulate αB-crystallin, which in turn could be acting in a structural role, or as a chaperone for the structural proteins of the myofibres.

# Chapter 2: The *prox* Gene Family

## 2.1 Introduction

We wished to elucidate the role of Prox1, the vertebrate homologue of the *Drosophila prospero* gene, in development by examining its role in zebrafish slow muscle formation. Given the duplication of the genome in the Actinopterygian lineage (Postlethwait et al., 1998), we decided to look for the presence of additional Prox genes in zebrafish. If present, these additional genes may be functionally redundant with Prox1 and could lead to a misinterpretation of Prox1 function in loss of function approaches. We searched for additional genes using both practical and computational methods. Southern analysis of zebrafish cDNA libraries isolated full-length sequences for *prox*1, and a novel *prox* gene, *prox2*. Computational analysis was carried out on the emerging zebrafish genome sequence as well as the two pufferfish genomes of *Fugu rubripes* and *Tetraodon nigiroviridus*. In addition to the genomic sequence of *prox1*, and the novel *prox2* gene, a third *prox* gene, *prox3* was predicted from the emerging genome sequence. A partial cDNA of *prox3* was identified in the EST database by identity to the predicted coding sequence. All three genes could be identified in the pufferfish genomes as well as a potential distantly related fourth member.

Due to the possible genome duplication, many mammalian genes have multiple homologues in zebrafish (Amores et al., 1998; Postlethwait et al., 1998). However, phylogenetic analysis of the novel Prox protein sequences suggested that the duplication of the Prox genes was ancestral and not a fish specific duplication. Examination of the mouse and human genomic sequences identified the presence of a second Prox gene in both species. In order to avoid confusion with the zebrafish genes we refer to this new mammalian Prox gene as *proxB*.

We show evidence in the genomic sequence flanking the *prox* genes for synteny between some of the newly identified members. This synteny information allowed the evolution of the Prox genes to be determined. We present evidence for the duplication of the *Drosophila prospero* gene in vertebrate evolution prior to the divergence of *Actinopterygii* and *Sarcopterygii*, followed by a second duplication in the Actinopterygian lineage. There have been two genome scale duplication events early in the vertebrate lineage (reviewed in Meyer and Schartl, 1999), and a further duplication in the Actinopterygian lineage. The observed duplications of the Prox genes are consistent with the loss of one of the two copies from the first round of duplication, and retention of both copies from the second vertebrate duplication and all four copies resulting from the Actinopterygian duplication.

We demonstrate that Prox1 is a member of a multigene family of *prospero* related transcription factors. We describe the discovery of this novel multigene family and examine its evolution. We also characterise the expression pattern of the newly identified zebrafish *prox2* and *prox3* genes.

## 2.2 Results

### 2.2.1 Isolation of *prox* cDNAs

A 15-19 hour polyA lambda phage zebrafish cDNA library was screened using a 500bp region of the *prox*1 homeodomain, labelled with $^{32}$P-CTP, as a probe. Two positive clones were identified and isolated. End sequencing and restriction mapping demonstrated one of the clones to be a full-length clone of *prox*1 (2.4kb). Sequencing of the other clone (3.5kb), showed it was not *prox*1, although it weakly bound the *prox*1 homeodomain region probe in a Southern hybridisation (Fig. 2.1). A zebrafish genomic cosmid used as a negative control did not bind the *prox*1 probe (Fig. 2.1).

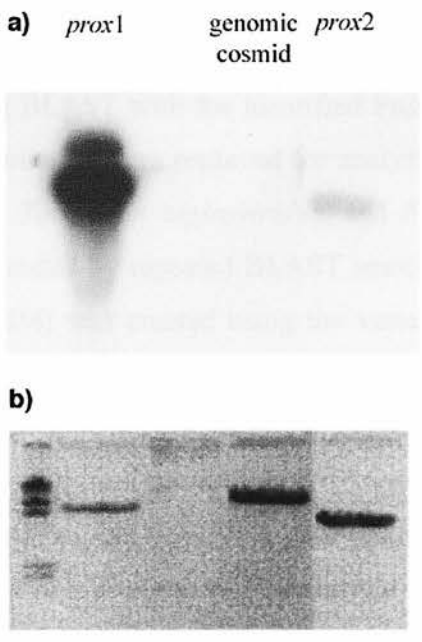



**Fig. 2.1 Southern hybridisation of *prox1* and *prox2***

a) Southern hybridisation of *prox1* and *prox2* cDNAs with the *prox1* homeodomain and *prospero* domain. Very strong hybridisation is observed with *prox1* and weak hybridisation to *prox2*. No hybridisation is observed to a genomic cosmid. b) Agarose gel showing the equal loading of the genomic cosmid and the *prox2* cDNA.

The novel cDNA is referred to hereafter as *prox2* and is characterised in section 2.2.3. The full cDNA sequence is given in Appendix 1.2.1.

### 2.2.2 Genomic Sequence Analysis

The zebrafish, *Tetraodon nigiroviridus*, *Fugu rubripes,* human, and mouse genomes were searched using BLAST with the identified Prox proteins. Genomic sequences matching the Prox proteins were retrieved for analysis. Matches to the unassembled genomes, zebrafish, *Tetraodon nigiroviridus*, and *Fugu rubripes*, were assembled into contigs and extended by repeated BLAST searching and assembling. A hidden Markov Model (HMM) was created using the vertebrate Prox1 sequences (human Prox1 Q92786, mouse Prox1 NM_032963, *X.leavis* Prox1 BAB17310, zebrafish Prox1 AAC70926, chick Prox1 Q91018) and zebrafish Prox2. This model was then used to predict the coding sequences of the Prox genes from the genomic sequences identified using genewise. The HMM was used in combination with a modified gene model to allow for the prediction of U12 introns. The modified gene model was created by M.S. Taylor.

Analysis of the zebrafish genome isolated a fragment of a third Prox gene, *prox*3. The predicted coding sequence for this gene was used to search the zebrafish EST database and a partial cDNA was identified (accession number BI886037). The expression and characterisation of zebrafish *prox*3 is described later (2.2.4). Analysis of the *Fugu* genome isolated homologues of zebrafish *prox1*, *prox2* and *prox3* as well as an additional possible family member, *prox4*. *prox4* has a homeodomain and prospero domain that can be identified by the HMM model, but the model is unable to predict any other exons. As a good prediction of the sequence could not be made it was not included in the subsequent analysis, but is discussed later (2.2.5)

In the human genome, a second Prox gene was identified at chromosome 14q24.3. Similarly in the mouse, a second gene was identified on chromosome 12. The human

and mouse genes do not appear to be homologues of zebrafish *prox2*, therefore, to avoid confusion, are referred to as mouse and human *proxB*.

The predicted protein and coding sequence for each of the genes isolated is given in the Appendix (Appendix 1.3).

### 2.2.3 Characterisation of *prox2*

Sequence analysis of *prox2* showed the presence of a homeodomain and *prospero* box at the 3' end of the cDNA. The clone appears to be full-length due to the presence of a start site and a polyA tail, identifying *prox2* as second member of the Prox family in the zebrafish.

We characterised the expression of *prox2* by in situ hybridisation. At the 8-somite stage expression was observed at the midbrain-hindbrain boundary (Fig. 2.2a) in the two lateral stripes in the mesoderm in the posterior of the embryo, in a pattern similar to haematopoetic precursors (Fig. 2.2b). The expression at the midbrain hindbrain boundary was clearly visible at the 13-somite and 18-somite stages (Fig.2.2c,d). At the 22-somite stage expression was only detectable in the intermediate cell mass, consisting of haematopoetic progenitors, in the tail, posterior to the limit of the yolk extension, the brain expression no longer detectable (Fig.2.2e). At 28-somites *prox2* was observed in the lens (Fig.2.2f), rhombomeres one to six (Fig.2.2g), the liver and the pancreas (Fig. 2.2h). The rhombomeric expression pattern is restricted to two cubes in each rhombomere, one either side of the midline, with a border of non-expressing cells surrounding each cube (Fig.2.2g). As described for *prox1*, the highest levels of expression were in the lens (Fig.2.2f). No expression was detected at 36-somites, or at any subsequent stages examined, and no expression was observed in the muscle at any stage.

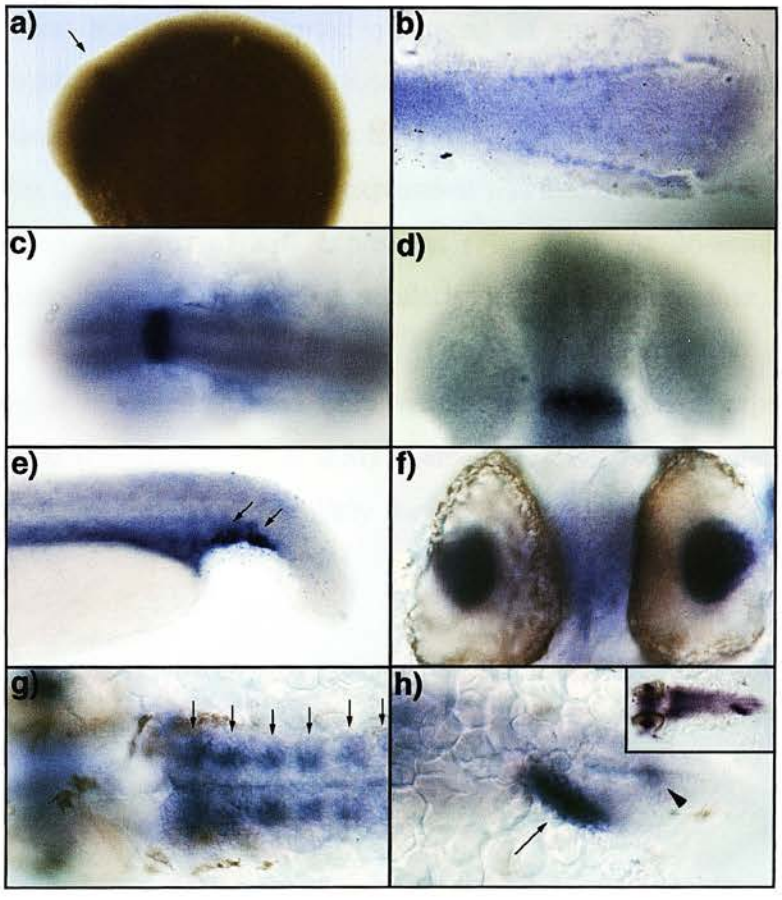



**Fig. 2.2 *prox*2 expression pattern**

In situ hybridisation using the full-length *prox*2 cDNA as a probe.
All views are dorsal and follow the convention anterior to the left unless otherwise stated.
a) Lateral view of an 8-somite embryo showing very weak expression in a midbrain hindbrain stripe
b) At the 8-somite stage expression is seen in the heamatopoetic precursors at the posterior of the embryo.
c) At 13-somites a midbrain-hindbrain stripe is clearly visible.
d) At 18-somites the midbrain-hindbrain stripe is still strongly expressing, dorsal view anterior to the top.
e) At the 22-somite stage expression is only visible in the intermediate cell mass (arrows).
f-h ) Expression at 28-somites, there is very strong lens expression (f), two patches of expression either side of the midline in rhombomeres one to six (g, arrows), and expression in the pancreas (arrowhead) and liver (arrow; f). The insert shows the position of the liver within the embryo.

### 2.2.4 Characterisation of *prox3*

The EST (identified in 2.2.2) was sequenced to confirm its identity and showed a significant match to the known Prox genes ($e10^{-16}$ by BLASTx). The EST is not full-length, an open reading frame is detected starting from the first base. The *prox3* EST showed homology to one of the predicted *Fugu* gene (57% identity and 70% similarity over the aligned region).

Expression analysis was carried out using an anti-sense probe to the EST and demonstrated expression at the 8-somite stage in a broad stripe, corresponding to rhombomeres four and five of the hindbrain, with higher expression in rhombomere five (Fig. 2.3a,b). Rhombomeric expression was also detected at the 16-somite stage, as well as the start of lens expression (Fig. 2.3c). At 26-somites rhombomeric expression is much weaker, particularly in rhombomere four, but strong lens expression is evident (Fig. 2.3d,e). At the 32-somite stage no rhombomere four expression is observed, and only weak rhombomere five expression, but lens expression is maintained (Fig. 2.3f). Lens expression was also detectable at the 36-somite stage (Fig. 2.3g). Similarly to *prox2*, expression was never detected in the muscle.

In the subsequent computational analyses the predicted *Fugu prox3* was used, as more sequence was available.

### 2.2.5 Characterisation of *prox4*

The predicted *Fugu prox4* sequence contained a homeodomain and *prospero* domain which showed 75% identity to the corresponding region in zebrafish Prox1. The genewise program also predicted the introns in the same location, as the other *prox* genes, but did not predict any sequence 5' of the homeodomain region. A sequence in the *Tetraodon* genome identified as a match to the Prox proteins by BLAST, was

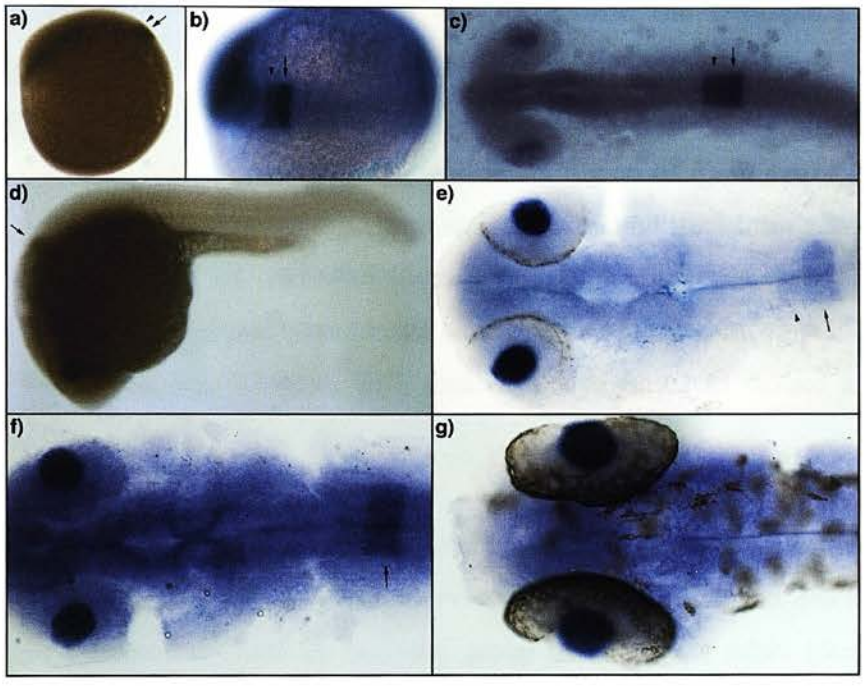



**Fig. 2.3 *prox3* expression pattern**

In situ hybridisation using a probe against *prox3*. All views are dorsal and follow the convention anterior to the left unless otherwise stated.
a,b) Lateral view (a) and dorsal view (b) showing expression in rhombomeres four (arrowhead) and five (arrow) at the 8-somite stage.
c) The rhombomeric expression pattern is maintained at 16-somites, and the first lens expression is observed.
d) Lateral view of a 26-somite embryos showing strong lens expression and rhombomere 5 expression.
e) Dorsal view of a 26-somite embryo showing that weak rhombomere four expression is just detectable (arrowhead).
f) At 32-somites strong lens expression is maintained and only rhombomere five continues to express *prox3*.
g) At 36 hours only lens expression is detected.

identified as being homologous to *Fugu prox4*. The *Tetraodon* sequence had 89.4% identity and 93.4% similarity at the amino acid level to the predicted HD and PD of *Fugu* Prox4. Genewise was unable to predict a protein for the *Tetraodon* sequence.

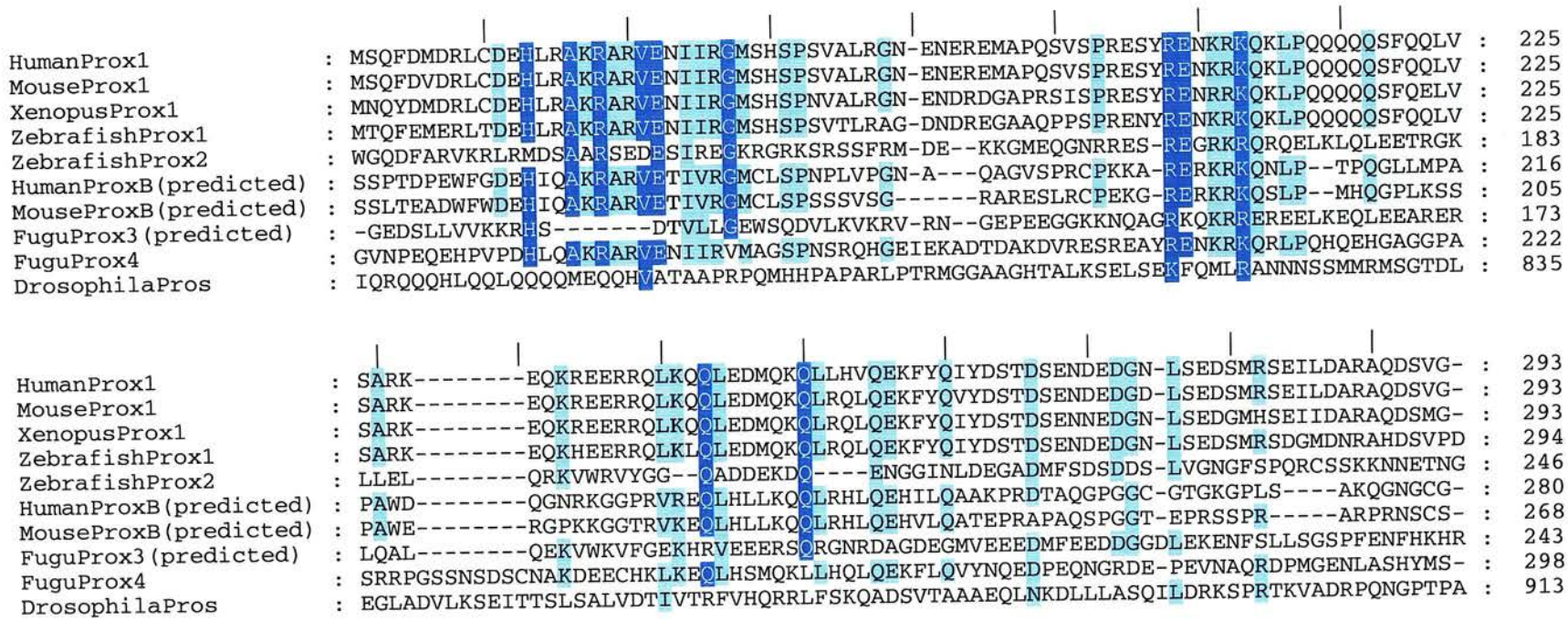
In order to reduce the stringency of the prox HMM the *Drosophila* Pros and *C.elegans* ceh-26, the *C.elegans* homologue of *prospero,* sequences were included in the HMM (Accession numbers P29617 and P34522). Using this model it was possible to predict exons 5' of the homeodomain region of *Fugu prox4* (Appendix 1.3.5). A region of the predicted coding sequence was amplified from genomic DNA and used as a probe on zebrafish liver and early somitogenesis grided libraries (RZPD). No positive clones were identified. Analysis of the zebrafish whole genome shotgun sequence and zebrafish EST databases also failed to isolate a zebrafish homologue of *Fugu prox*4.

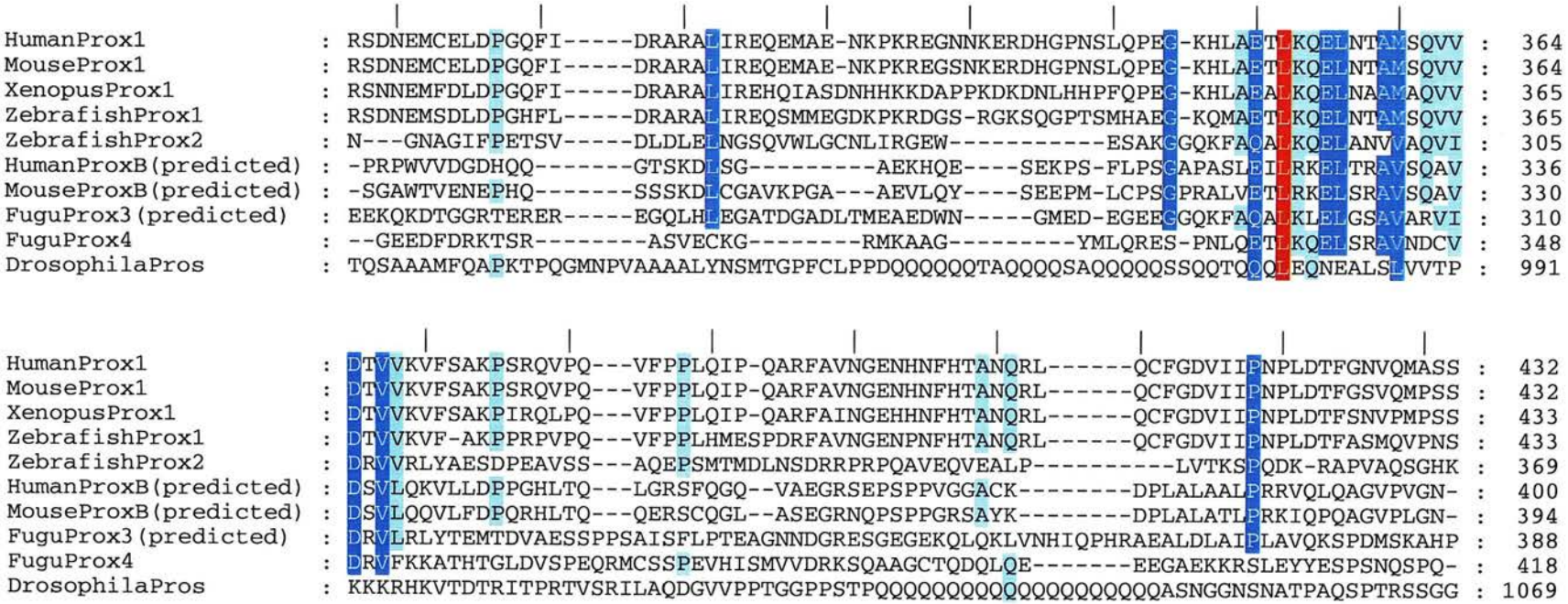
As a good prediction of the Prox4 protein sequence could not be made it was excluded from subsequent analysis. However, from the sequence that was available it would be expected to form an outlying group to all of the other identified and predicted Prox proteins.

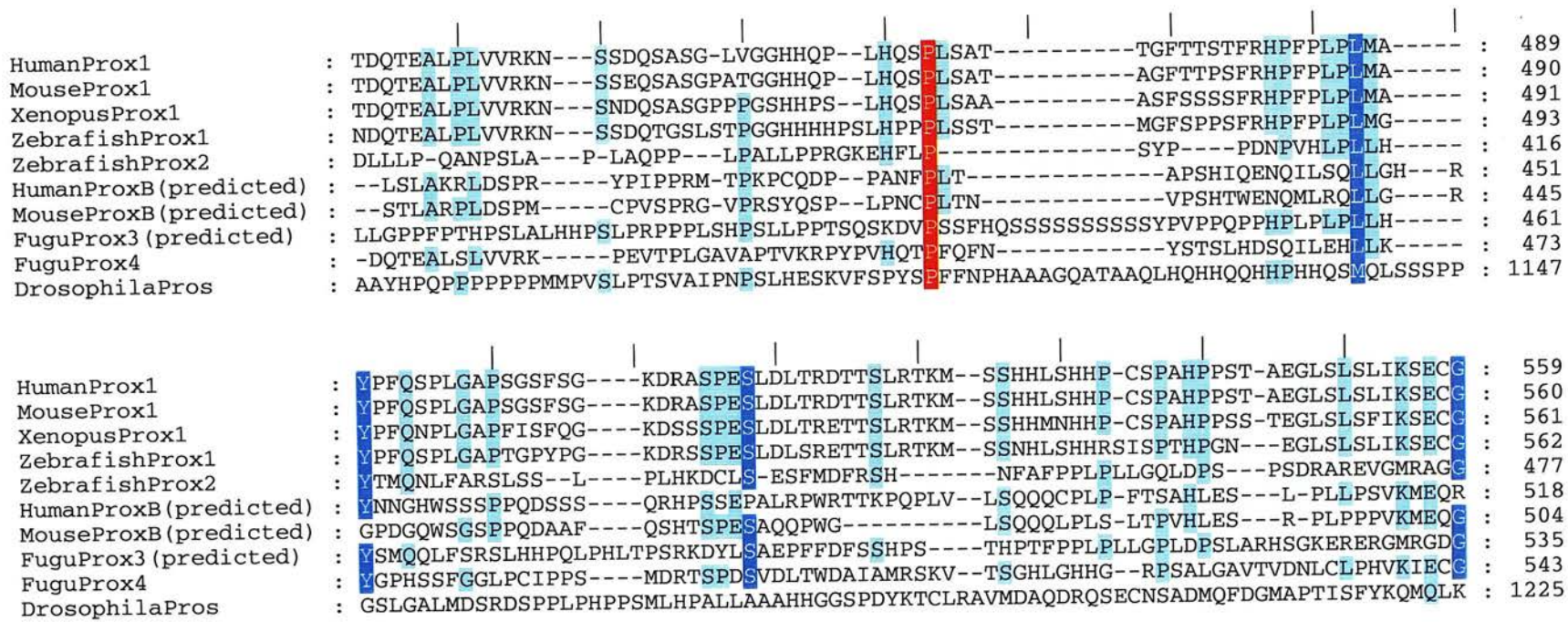
### 2.2.6 Sequence Comparison of Prox Proteins

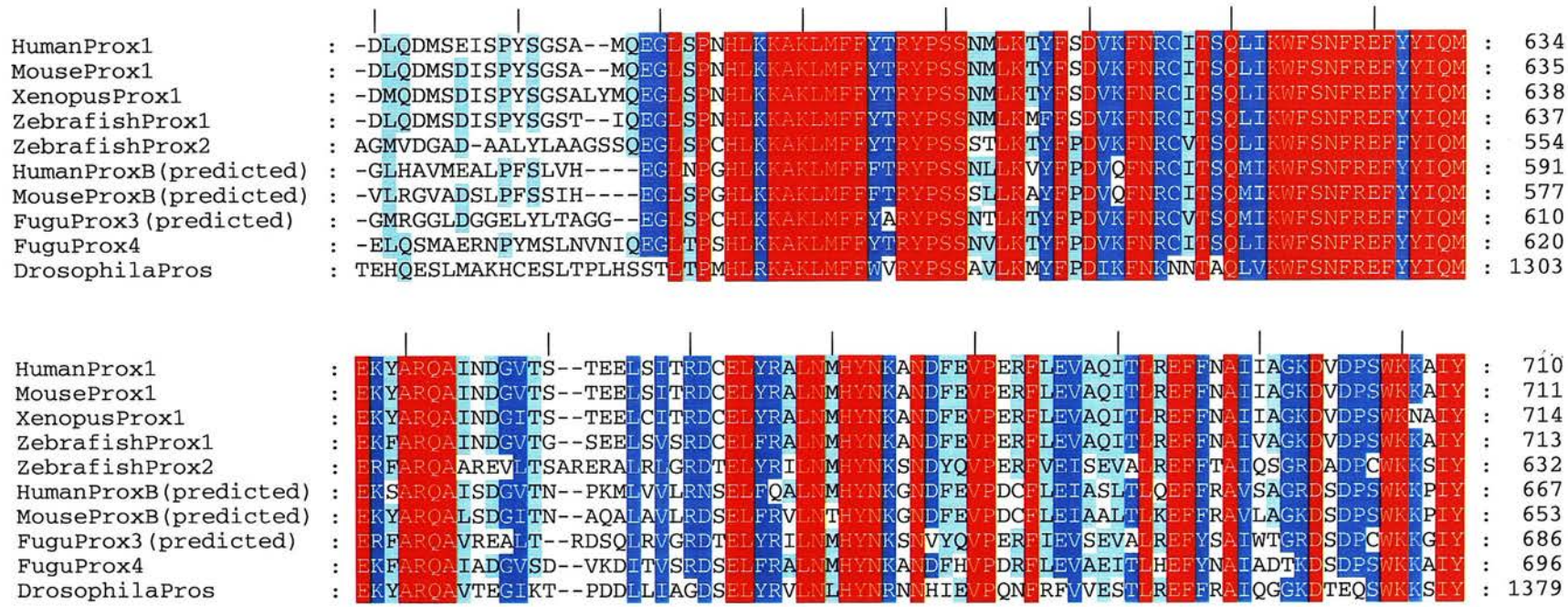
An alignment of Prox1 protein sequences and the newly identified Prox proteins was created using ClustalW (Fig 2.4). The N-terminal region has very little conservation, with the exception of three short regions of conservation, none of which are conserved with Pros. The C-terminal homeodomain and *prospero* domain is highly conserved, as demonstrated by the high level of similar and identical residues (Fig. 2.4). Pairwise analysis of conservation was carried out using needle, for both the entire protein and for the C-terminal homeodomain and *prospero* domains (Fig. 2.5). The values given are overall identity and similarity, this is the values over the total length of the longest sequence, not just the values for which the sequences align.


HumanProx1 : -MPDHDSTALLSRQTKRRRVDIGVKRTVGTASAFFAKARATFFSAM--NPQGSEQDVEYSVVQHADGEKSNVLRKLLK : 75
MouseProx1 : -MPDHDSTALLSRQTKRRRVDIGVKRTVGTASAFFAKARATFFSAM--NPQGSEQDVEYSVVQHADGEKSNVLRKLLK : 75
XenopusProx1 : -MPDHDSTGLLSRQTKRRRVDIGVKRTVGTASAFFAKASATFFSAM--NPQGSEQDVEYSVVQHADGEKSNVLRKLLK : 75
ZebrafishProx1 : -MPDHDSTSLLSRQTKRRRVDIGVKRTVGQASVVFTRAKATFLSAM--NPHGSEQDVECSVVQHADGEKSNVLRKLLK : 75
ZebrafishProx2 : ----------MDSPSD----RFHQQPSQMCPFELYRSLPDSNPSGHR-----FPPGSLSSHPLFSP : 47
HumanProxB(predicted) : QTESITASCRLT-QIKGCGQHAGLSTEV---SCFSAAQRQSFHDCMGVNISGYPATVESGWISSFWPQDISVHCCFIK : 74
MouseProxB(predicted) : -MPRADITGRPT-NIRRARLRVIGSRNK---SALPVKGSITLAFLNPGAVDVHSLRVKK------TLQHALAHLSLLN : 67
FuguProx3(predicted) : ----------FAVMDSPSD----LFEDSGRQIHTFAS--TLSSSDLTGVQ-----LQPSTRHPPPAFRP : 48
FuguProx4 : -LPQHLSLRRPS---ARRLFSAQQVRSHLSAAVFALHPSEQFFSEM--NLSHPDQNMHGAGDGCLEDDKPDVMFPCFR : 72
DrosophilaPros : LSPSPTLTGDGDVSPNHKEETGQERPGSSSPSPSPLKPKTSLGESSDSGANMLSQMMSKMMSGKLHNPLVGVGHPALP : 679
HumanProx1 : RANS--YEDAMMPFPGATIISQL-LKNNMNKN-GGTEPSFQASGLSSTGSEVHQEDICSNSSRDSPPECLSP-FGRPT : 148
MouseProx1 : RANS--YEDAMMPFPGATIISQL-LKNNMNKN-GGTEPSFQASGLSSTGSEVHQEDICSNSSRDSPPECLSP-FGRPT : 148
XenopusProx1 : RANS--YEDAMMPFPGATIISQL-LKNNMNKN-CGTEPSFQASGLSSAGSEIHQEDACINSSRDSPQECLSP-FGRPT : 148
ZebrafishProx1 : RANS--YEDAVMPFPGATIISQL-LKNNLTKN-GGSEPNFQGSGLSSTGSEMQQEDACSNSSRGSPQECLSP-FSRPS : 148
ZebrafishProx2 : LMHP----NSMSQKWGSFRQRQG-LFPGLLVDDDVSEEDVFGKREFGATSSSQRGSGERGQAPEP----------SD : 109
HumanProxB(predicted) : HTGP--ICYPSRSVFWSTQTTRN-LRNNTLNNPQGRFSSFSGSMLFSSMTLSTRLSFCLQLDRDSP---FPW-SQVPS : 145
MouseProxB(predicted) : LAG---LSKALGSSWELCMSIRR-GRTWIHLLLSCFLLSLGPAPTWQTCMDQERSPATAEAGRDS----FPS-GQLPS : 136
FuguProx3(predicted) : LGFP--LIHNLLHPGGATRRVGGQLNPNLRSHRHLEEVN----LEENMQERDGRVDHEMDGLIE---------- : 106
FuguProx4 : RNV---YDESLSSYSSGSIISQL-LRKTIHSARALDESHFY--LSSSGGADCSQEDQNSVSSKDSTMEATSPGAHVST : 144
DrosophilaPros : QGFPPLLQHMGDMSHAAAMYQQFFFEQEARMAKEAAEQQQQQQQQQQQQQQQQQQQEQQRRFEQEQQEQQRRKEEQQQQ : 757


HumanProx1 : MSQFDMDRLCDEHLRAKRARVENIIRCMSHSPSVALRGN-ENEREMAPQSVSPRESYRENKRKQKLPQQQQQSFQQLV : 225
MouseProx1 : MSQFDVDRLCDEHLRAKRARVENIIRCMSHSPSVALRGN-ENEREMAPQSVSPRESYRENKRKQKLPQQQQQSFQQLV : 225
XenopusProx1 : MNQYDMDRLCDEHLRAKRARVENIIRCMSHSPNVALRGN-ENDRDGAPRSISPRESYRENRRKQKLPQQQQQSFQELV : 225
ZebrafishProx1 : MTQFEMERLTDEHLRAKRARVENIIRCMSHSPSVTLRAG-DNDREGAAQPPSPRENYRENKRKQKLPQQQQQSFQQLV : 225
ZebrafishProx2 : WGQDFARVKRLRMDSAARSEDESIRECKRGRKSRSSFRM-DE--KKGMEQGNRRES-REGRKRQRQELKLQLEETRGK : 183
HumanProxB(predicted) : SSPTDPEWFGDEHIQAKRARVETIVRCMCLSPNPLVPGN-A---QAGVSPRCPKKA-RERKRKQNLP--TPQGLLMPA : 216
MouseProxB(predicted) : SSLTEADWFWDEHIQAKRARVETIVRCMCLSPSSSVSG------RARESLRCPEKG-RERKRKQSLP--MHQGPLKSS : 205
FuguProx3(predicted) : -GEDSLLVVKKRHS-------DTVLLCEWSQDVLKVKRV-RN--GEPEEGGKKNQAGRKQKRREREELKEQLEEARER : 173
FuguProx4 : GVNPEQEHPVPDHLQAKRARVENIIRVMAGSPNSRQHGEIEKADTDAKDVRESREAYRENKRKQRLPQHQEHGAGGPA : 222
DrosophilaPros : IQRQQQHLQQLQQQQMEQQHVATAAPRPQMHHPAPARLPTRMGGAAGHTALKSELSEKFQMLRANNNSSMMRMSGTDL : 835
HumanProx1 : SARK--------EQKREERRQLKQQLEDMQKQLLHVQEKFYQIYDSTDSENDEDGN-LSEDSMRSEILDARAQDSVG- : 293
MouseProx1 : SARK--------EQKREERRQLKQQLEDMQKQLRQLQEKFYQVYDSTDSENDEDGD-LSEDSMRSEILDARAQDSVG- : 293
XenopusProx1 : SARK--------EQKREERRQLKQQLEDMQKQLRQLQEKFYQIYDSTDSENNEDGN-LSEDGMHSEIIDARAQDSMG- : 293
ZebrafishProx1 : SARK--------EQKHEERRQLKLQLEDMQKQLRQLQEKFYQIYDSTDSENDEDGN-LSEDSMRSDGMDNRAHDSVPD : 294
ZebrafishProx2 : LLEL--------QRKVWRVYGG--QADDEKDQ----ENGGINLDEGADMFSDSDDS-LVGNGFSPQRCSSKKNNETNG : 246
HumanProxB(predicted) : PAWD--------QGNRKGGPRVREQLHLLKQQLRHLQEHILQAAKPRDTAQGPGGC-GTGKGPLS----AKQGNGCG- : 280
MouseProxB(predicted) : PAWE--------RGPKKGGTRVKEQLHLLKQQLRHLQEHVLQATEPRAPAQSPGGT-EPRSSPR-----ARPRNSCS- : 268
FuguProx3(predicted) : LQAL--------QEKVWKVFGEKHRVEEERSQRGNRDAGDEGMVEEEDMFEEDDGGDLEKENFSLLSGSPFENFHKHR : 243
FuguProx4 : SRRPGSSNSDSCNAKDEECHKLKEQLHSMQKLLHQLQEKFLQVYNQEDPEQNGRDE-PEVNAQRDPMGENLASHYMS- : 298
DrosophilaPros : EGLADVLKSEITTSLSALVDTIVTRFVHQRRLFSKQADSVTAAAEQLNKDLLLASQILDRKSPRTKVADRPQNGPTPA : 913


HumanProx1 : RSDNEMCELDPGQFI-----DRARALIREQEMAE-NKPKREGNNKERDHGPNSLQPEG-KHLAETLKQELNTAMSQVV : 364
MouseProx1 : RSDNEMCELDPGQFI-----DRARALIREQEMAE-NKPKREGSNKERDHGPNSLQPEG-KHLAETLKQELNTAMSQVV : 364
XenopusProx1 : RSNNEMFDLDPGQFI-----DRARALIREHQIASDNHHKKDAPPKDKDNLHHPFQPEG-KHLAEALKQELNAAMAQVV : 365
ZebrafishProx1 : RSDNEMSDLDPGHFL-----DRARALIREQSMMEGDKPKRDGS-RGKSQGPTSMHAEG-KQMAETLKQELNTAMSQVV : 365
ZebrafishProx2 : N---GNAGIFPETSV-----DLDLELNGSQVWLGCNLIRGEW----------ESAKGGQKFAQALKQELANVVAQVI : 305
HumanProxB(predicted) : -PRPWVVDGDHQQ-------GTSKDLSG---------AEKHQE----SEKPS-FLPSGAPASLEILRKELTRAVSQAV : 336
MouseProxB(predicted) : -SGAWTVENEPHQ-------SSSKDLCGAVKPGA---AEVLQY----SEEPM-LCPSGPRALVETLRKELSRAVSQAV : 330
FuguProx3(predicted) : EEKQKDTGGRTERER-----EGQLHLEGATDGADLTMEAEDWN-----GMED-EGEEGGQKFAQALKLELGSAVARVI : 310
FuguProx4 : --GEEDFDRKTSR--------ASVECKG--------RMKAAG---------YMLQRES-PNLQETLKQELSRAVNDCV : 348
DrosophilaPros : TQSAAAMFQAPKTPQGMNPVAAAALYNSMTGPFCLPPDQQQQQQTAQQQQSAQQQQQSSQQTQQQLEQNEALSLVVTP : 991
HumanProx1 : DTVVKVFSAKPSRQVPQ---VFPPLQIP-QARFAVNGENHNFHTANQRL------QCFGDVIIPNPLDTFGNVQMASS : 432
MouseProx1 : DTVVKVFSAKPSRQVPQ---VFPPLQIP-QARFAVNGENHNFHTANQRL------QCFGDVIIPNPLDTFGSVQMPSS : 432
XenopusProx1 : DTVVKVFSAKPIRQLPQ---VFPPLQIP-QARFAINGEHHNFHTANQRL------QCFGDVIIPNPLDTFSNVPMPSS : 433
ZebrafishProx1 : DTVVKVF-AKPPRPVPQ---VFPPLHMESPDRFAVNGENPNFHTANQRL------QCFGDVIIPNPLDTFASMQVPNS : 433
ZebrafishProx2 : DRVVRLYAESDPEAVSS---AQEPSMTMDLNSDRRPRPQAVEQVEALP----------LVTKSPQDK-RAPVAQSGHK : 369
HumanProxB(predicted) : DSVLQKVLLDPPGHLTQ---LGRSFQGQ--VAEGRSEPSPPVGGACK--------DPLALAALPRRVQLQAGVPVGN- : 400
MouseProxB(predicted) : DSVLQQVLFDPQRHLTQ---QERSCQGL--ASEGRNQPSPPGRSAYK--------DPLALATLPRKIQPQAGVPLGN- : 394
FuguProx3(predicted) : DRVLRLYTEMTDVAESSPPSAISFLPTEAGNNDGRESGEGEKQLQKLVNHIQPHRAEALDLAIPLAVQKSPDMSKAHP : 388
FuguProx4 : DRVFKKATHTGLDVSPEQRMCSSPEVHISMVVDRKSQAAGCTQDQLQE-------EEGAEKKRSLEYYESPSNQSPQ- : 418
DrosophilaPros : KKKRHKVTDTRITPRTVSRILAQDGVVPPTGGPPSTPQQQQQQQQQQQQQQQQQQQQQQASNGGNSNATPAQSPTRSSGG : 1069


HumanProx1 : TDQTEALPLVVRKN---SSDQSASG-LVGGHHQP--LHQSPLSAT----------TGFTTSTFRHPFPLPLMA----- : 489
MouseProx1 : TDQTEALPLVVRKN---SSEQSASGPATGGHHQP--LHQSPLSAT----------AGFTTPSFRHPFPLPLMA----- : 490
XenopusProx1 : TDQTEALPLVVRKN---SNDQSASGPPPGSHHPS--LHQSPLSAA----------ASFSSSSFRHPFPLPLMA----- : 491
ZebrafishProx1 : NDQTEALPLVVRKN---SSDQTGSLSTPGGHHHHPSLHPPPLSST----------MGFSPPSFRHPFPLPLMG----- : 493
ZebrafishProx2 : DLLLP-QANPSLA---P-LAQPP---LPALLPPRGKEHFLP----------SYP----PDNPVHLPLLH----- : 416
HumanProxB(predicted) : --LSLAKRLDSPR-----YPIPPRM-TPKPCQDP--PANFPLT----------APSHIQENQILSQLLGH---R : 451
MouseProxB(predicted) : --STLARPLDSPM-----CPVSPRG-VPRSYQSP--LPNCPLTN----------VPSHTWENQMLRQLLG----R : 445
FuguProx3(predicted) : LLGPPFPTHPSLALHHPSLPRPPPLSHPSLLPPTSQSKDVPSSFHQSSSSSSSSSSYPVPPQPPHPLPLPLLH----- : 461
FuguProx4 : -DQTEALSLVVRK-----PEVTPLGAVAPTVKRPYPVHQTPFQFN----------YSTSLHDSQILEHLLK----- : 473
DrosophilaPros : AAYHPQPPPPPPPMMPVSLPTSVAIPNPSLHESKVFSPYSPFFNPHAAAGQATAAQLHQHHQQHHPHHQSMQLSSSPP : 1147
HumanProx1 : YPFQSPLGAPSGSFSG----KDRASPESLDLTRDTTSLRTKM--SSHHLSHHP-CSPAHPPST-AEGLSLSLIKSECG : 559
MouseProx1 : YPFQSPLGAPSGSFSG----KDRASPESLDLTRDTTSLRTKM--SSHHLSHHP-CSPAHPPST-AEGLSLSLIKSECG : 560
XenopusProx1 : YPFQNPLGAPFISFQG----KDSSSPESLDLTRETTSLRTKM--SSHHMNHHP-CSPAHPPSS-TEGLSLSFIKSECG : 561
ZebrafishProx1 : YPFQSPLGAPTGPYPG----KDRSSPESLDLSRETTSLRTKM--SSNHLSHHRSISPTHPGN---EGLSLSLIKSECG : 562
ZebrafishProx2 : YTMQNLFARSLSS--L----PLHKDCLS-ESFMDFRSH-------NFAFPPLPLLGQLDPS---PSDRAREVGMRAGG : 477
HumanProxB(predicted) : YNNGHWSSSPPQDSSS----QRHPSSEPALRPWRTTKPQPLV--LSQQQCPLP-FTSAHLES---L-PLLPSVKMEQR : 518
MouseProxB(predicted) : GPDGQWSGSPPQDAAF----QSHTSPESAQQPWG----------LSQQQLPLS-LTPVHLES---R-PLPPPVKMEQG : 504
FuguProx3(predicted) : YSMQQLFSRSLHHPQLPHLTPSRKDYLSAEPFFDFSSHPS----THPTFPPLPLLGPLDPSLARHSGKERERGMRGDG : 535
FuguProx4 : YGPHSSFGGLPCIPPS----MDRTSPDSVDLTWDAIAMRSKV--TSGHLGHHG--RPSALGAVTVDNLCLPHVKIECG : 543
DrosophilaPros : GSLGALMDSRDSPPLPHPPSMLHPALLAAAHHGGSPDYKTCLRAVMDAQDRQSECNSADMQFDGMAPTISFYKQMQLK : 1225


HumanProx1 : -DLQDMSEISPYSGSA--MQEGLSPNHLKKAKLMFFYTRYPSSNMLKTYFSDVKFNRCITSQLIKWFSNFREFYYIQM : 634
MouseProx1 : -DLQDMSDISPYSGSA--MQEGLSPNHLKKAKLMFFYTRYPSSNMLKTYFSDVKFNRCITSQLIKWFSNFREFYYIQM : 635
XenopusProx1 : -DMQDMSDISPYSGSALYMQEGLSPNHLKKAKLMFFYTRYPSSNMLKTYFSDVKFNRCITSQLIKWFSNFREFYYIQM : 638
ZebrafishProx1 : -DLQDMSDISPYSGST--IQEGLSPNHLKKAKLMFFYTRYPSSNMLKMFFSDVKFNRCITSQLIKWFSNFREFYYIQM : 637
ZebrafishProx2 : AGMVDGAD-AALYLAAGSSQEGLSPCHLKKAKLMFFYTRYPSSSTLKTYFPDVKFNRCVTSQLIKWFSNFREFFYIQM : 554
HumanProxB(predicted) : -GLHAVMEALPFSLVH----EGLNPGHLKKAKLMFFFTRYPSSNLLKVYFPDVQFNRCITSQMIKWFSNFREFYYIQM : 591
MouseProxB(predicted) : -VLRGVADSLPFSSIH----EGLSPGHLKKAKLMFFFTRYPSSSLLKAYFPDVQFNRCITSQMIKWFSNFREFYYIQM : 577
FuguProx3(predicted) : -GMRGGLDGGELYLTAGG--EGLSPCHLKKAKLMFFYARYPSSNTLKTYFPDVKFNRCVTSQMIKWFSNFREFFYIQM : 610
FuguProx4 : -ELQSMAERNPYMSLNVNIQEGLTPSHLKKAKLMFFYTRYPSSNVLKTYFPDVKFNRCITSQLIKWFSNFREFYYIQM : 620
DrosophilaPros : TEHQESLMAKHCESLTPLHSSTLTPMHLRKAKLMFFWVRYPSSAVLKMYFPDIKFNKNNDAQLVKWFSNFREFYYIQM : 1303
HumanProx1 : EKYARQAINDGVTS--TEELSITRDCELYRALNMHYNKANDFEVPERFLEVAQITLREFFNAIIAGKDVDPSWKKAIY : 710
MouseProx1 : EKYARQAINDGVTS--TEELSITRDCELYRALNMHYNKANDFEVPERFLEVAQITLREFFNAIIAGKDVDPSWKKAIY : 711
XenopusProx1 : EKYARQAINDGITS--TEELCITRDCELYRALNMHYNKANDFEVPERFLEVAQITLREFFNAIIAGKDVDPSWKNAIY : 714
ZebrafishProx1 : EKFARQAINDGVTG--SEELSVSRDCELFRALNMHYNKANDFEVPERFLEVAQITLREFFNAIVAGKDVDPSWKKAIY : 713
ZebrafishProx2 : ERFARQAAREVLTSARERALRLGRDTELYRILNMHYNKSNDYQVPERFVEISEVALREFFTAIQSGRDADPCWKKSIY : 632
HumanProxB(predicted) : EKSARQAISDGVTN--PKMLVVLRNSELFQALNMHYNKGNDFEVPDCFLEIASLTLQEFFRAVSAGRDSDPSWKKPIY : 667
MouseProxB(predicted) : EKYARQALSDGITN--AQALAVLRDSELFRVLNTHYNKGNDFEVPDCFLEIAALTLKEFFRAVLAGKDSDPSWKKPIY : 653
FuguProx3(predicted) : ERFARQAVREALT--RDSQLRVGRDTELYRILNMHYNKSNVYQVPERFIEVSEVALREFYSAIWTGRDSDPCWKKGIY : 686
FuguProx4 : EKFARQAIADGVSD--VKDITVSRDSELFRALNMHYNKANDFHVPDRFLEVAEITLHEFYNAIADTKDSDPSWKKAIY : 696
DrosophilaPros : EKYARQAVTEGIKT--PDDLLIAGDSELYRVLNLHYNRNNHIEVPQNFRFVVESTLREFFRAIQGGKDTEQSWKKSIY : 1379

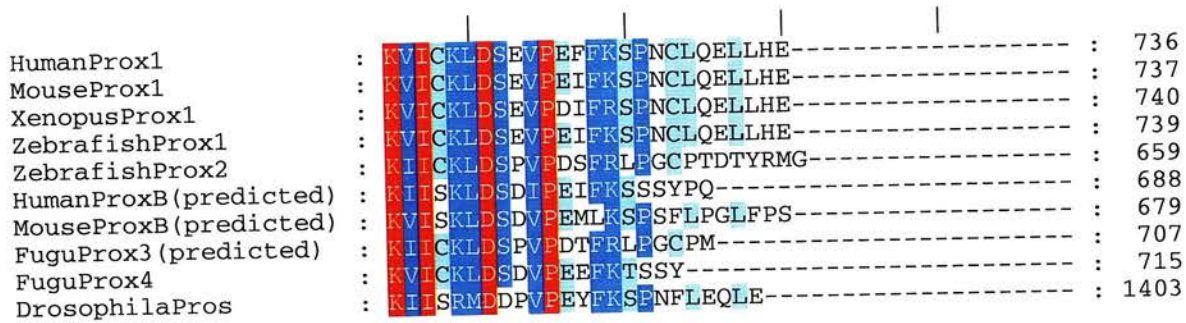



**Fig. 2.4 Alignment of Prox proteins**

Conserved identical amino acids are shown in red, 80% conserved identity and 100% similarities in dark blue, and 60% identity, 80% similarity in light blue. The homeodomain and *prospero* domains are highly conserved between al Prox proteins and Pros. The N-terminal region is largely unconserved with the exception of three short motifs at amino acids; 159-180, 205-214, and 350-368 (positions are given from Human Prox1). These regions are not conserved with Pros and are not known motifs, although their conservation between all Prox proteins suggests they may have a specific function.

| entire protein / HD and PD | Human Prox1 | Mouse Prox1 | Xenopus Prox1 | Zebrafish Prox1 | Zebrafish Prox2 | Fugu Prox2 (predicted) | Human ProxB (predicted) | Mouse ProxB (predicted) | Fugu Prox3 (predicted) | Drosophila Pros |
|---|---|---|---|---|---|---|---|---|---|---|
| Human Prox1 | | 98 (99) | 87 (93) | 82 (90) | 32 (47) | 32 (48) | 35 (50) | 37 (51) | 43 (57) | 17 (26) |
| Mouse Prox1 | 99 (99) | | 88 (94) | 83 (90) | 34 (48) | 33 (48) | 36 (50) | 38 (51) | 43 (57) | 18 (26) |
| Xenopus Prox1 | 95 (96) | 95 (97) | | 78 (86) | 33 (48) | 33 (48) | 34 (50) | 36 (50) | 41 (56) | 18 (27) |
| Zebrafish Prox1 | 93 (98) | 93 (98) | 89 (96) | | 32 (50) | 33 (47) | 35 (50) | 38 (51) | 42 (56) | 17 (26) |
| Zebrafish Prox2 | 66 (80) | 66 (80) | 67 (80) | 64 (80) | | 47 (58) | 32 (45) | 35 (49) | 32 (47) | 15 (22) |
| Fugu Prox2 (predicted) | 66 (80) | 66 (80) | 65 (78) | 64 (79) | 82 (88) | | 32 (45) | 33 (46) | 32 (44) | 15 (25) |
| Human ProxB (predicted) | 69 (82) | 69 (83) | 66 (81) | 70 (83) | 58 (76) | 61 (78) | | 56 (66) | 37 (50) | 14 (21) |
| Mouse ProxB (predicted) | 71 (83) | 71 (84) | 68 (82) | 71 (84) | 61 (79) | 61 (78) | 81 (89) | | 37 (51) | 16 (25) |
| Fugu Prox3 (predicted) | 75 (87) | 75 (87) | 72 (85) | 77 (86) | 62 (80) | 65 (80) | 73 (85) | 69 (81) | | 17 (26) |
| Drosophila Pros | 60 (75) | 60 (75) | 58 (75) | 58 (75) | 64 (71) | 53 (70) | 54 (74) | 59 (75) | 55 (71) | |

**Fig. 2.5 Amino acid identity and similarity between Prox proteins**

All of the vertebrate Prox proteins were compared to each other and to *Drosophila* Pros. The figures shown give the percentage identity with the similarity in brackets for the entire protein, top right, or the homeodomain (HD) and prospero domain (PD), bottom left.

Conservation is very high between the Prox1 proteins in all species examined, averaging 86% identity. Conservation was only slightly higher between the ProxB sequences (56% identity) and also the Prox2 sequences (47% identity) than to other Prox proteins. There is much less conservation between the different Prox proteins averaging 35% identity between groups. The low identity to Pros reflects the great difference in length between the Prox sequences and Pros, and in this case conservation of the homeodomain and *prospero* domain may be a better indicator of conservation.

Overall conservation of the C-terminal domains was very high, between Pros and the Prox proteins there is at least 53% identity and 70% similarity. Between the different Prox proteins, conservation is even higher with at least 58% identity and 76% similarity, although again conservation is higher still within groups (Fig. 2.5).

### 2.2.7 Genomic Organisation

Using the genomic sequence and genewise analysis it was possible to examine the genomic organisation of the identified genes. The number, size, position, and phase of introns was examined (Fig. 2.6). The positions given are those generated by genewise, with a couple of exceptions. In the case of mouse *proxB* there are additional exons, exons two, three, and four in the *proxB* prediction have introns of only one and 5 bases between them. This is probably due to a missing or additional base in the genomic sequence and these exons were joined. In the case of zebrafish Prox2 the genomic sequence had a gap in the assembly corresponding to the end of exon1. The size of the gap had been estimated in the assembly to be less than the length of coding sequence missing in the genomic sequence. It is therefore likely that there is no intron present in the missing genomic sequence, and this is assumed to be correct in Fig. 2.6.

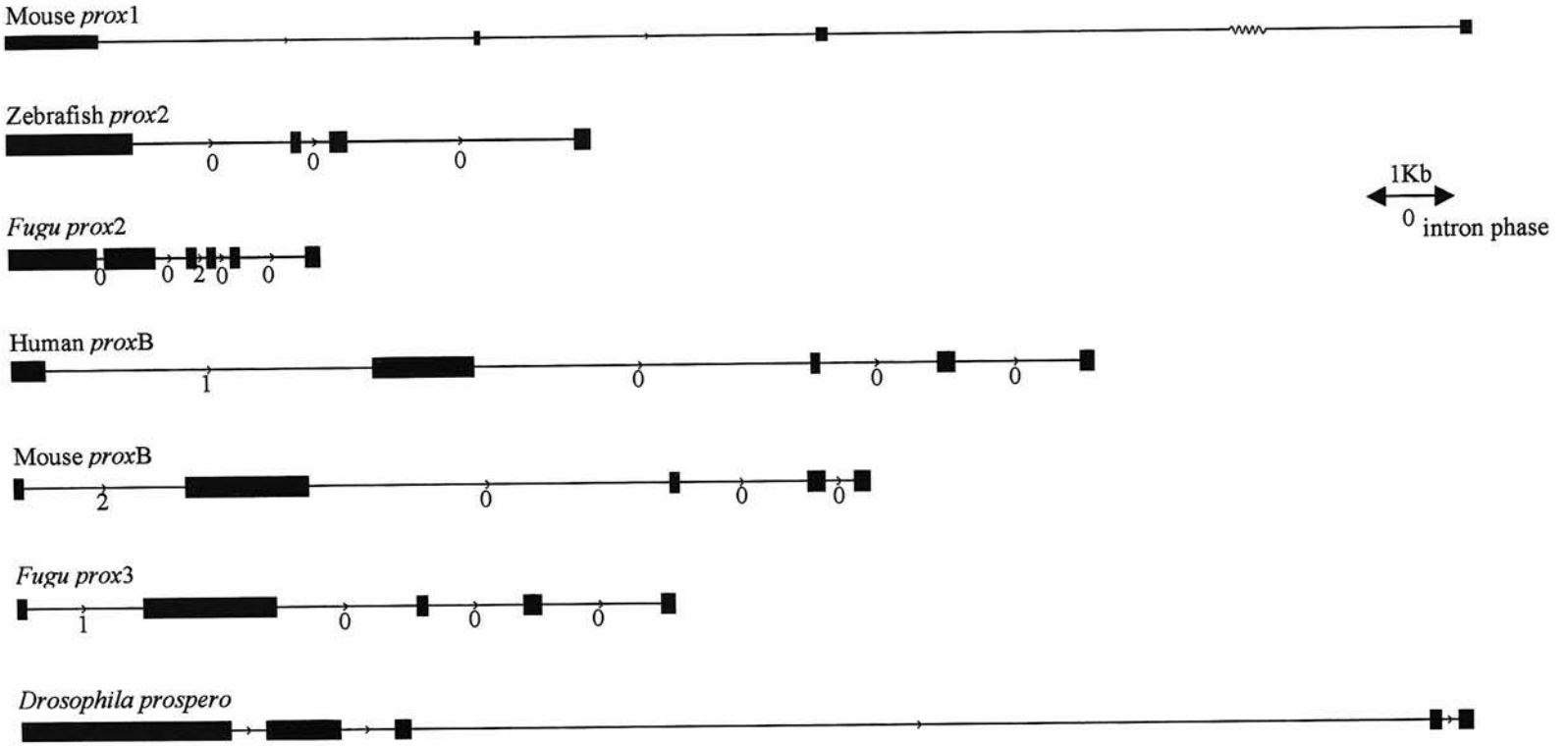



**Fig. 2.6 Gene Structure of the predicted Prox genes**

The genomic organisation of the Prox genes as described for mouse *prox*1 and *prospero*, and as identified for the novel Prox genes. The organisation of the human *prox* 1 gene is not shown, as it is very similar to mouse *prox*1, only differing slightly in intron size. The three terminal exons appear to be conserved in all of the *prox* genes, except *Fugu prox*2 where the penultimate exon has split into two. It is also clear that human *prox*B, mouse *prox*B, and *Fugu prox*3 have the same genomic organisation. This suggests *prox*3 ,may be most closely related to *prox*B. Untranslated exons reported in the literature are not shown, as they could not be predicted for the novel genes. The gene structure was drawn using ExonPlot (M.S Taylor, unpublished).

Exon number refers only to those in the coding sequence. In the Prox1 genomic sequences for mouse and human, 5' untranslated exons were identified, but these can not be predicted in the new sequences and therefore are not shown.

All of the Prox genes share the three exons that code for the homeodomain and *prospero* domain, with the exception of F*ugu prox2*. In the case of *Fugu prox2* it appears that the penultimate exon has had an intron insertion, after the divergence of zebrafish and *Fugu*. The *Fugu prox2* gene also differs from zebrafish *prox2* by the insertion of an additional intron in the first exon.

Both of the *proxB* genes and *Fugu prox3* share the same genomic structure, having an addition intron in the region corresponding to *prox1* intron1. This suggests that *Fugu prox3* may be more closely related to *proxB* than *prox1*.

### 2.2.8 Conservation of the *prospero* U12 Intron

*prospero* contains an unusual U12 spliced intron as part of a twintron structure. We examined the newly identified genes and Prox1 genes for the presence of the U12 intron and possible twintron structures. The presence of the U12 splice site was determined by the presence of the 5' and 3' splice sites and the 3' upstream element as identified by Hall and Padgett (1994; Fig. 2.7).

No ESTs could be identified in the available databases which had an insertion corresponding to an alternative splice as found in Drosophila. The intron sequence for zebrafish *prox2*, for which both genomic and cDNA sequence was available, was examined for open reading frames that continued into the intron. The sequence was then examined using StrataSplice for the presence of potential splice sites that would indicate the presence of a U2 intron. Several potential splice sites were identified that would could indicate the presence of a U2 spliced intron nested within the U12 sites (Fig. 2.8).

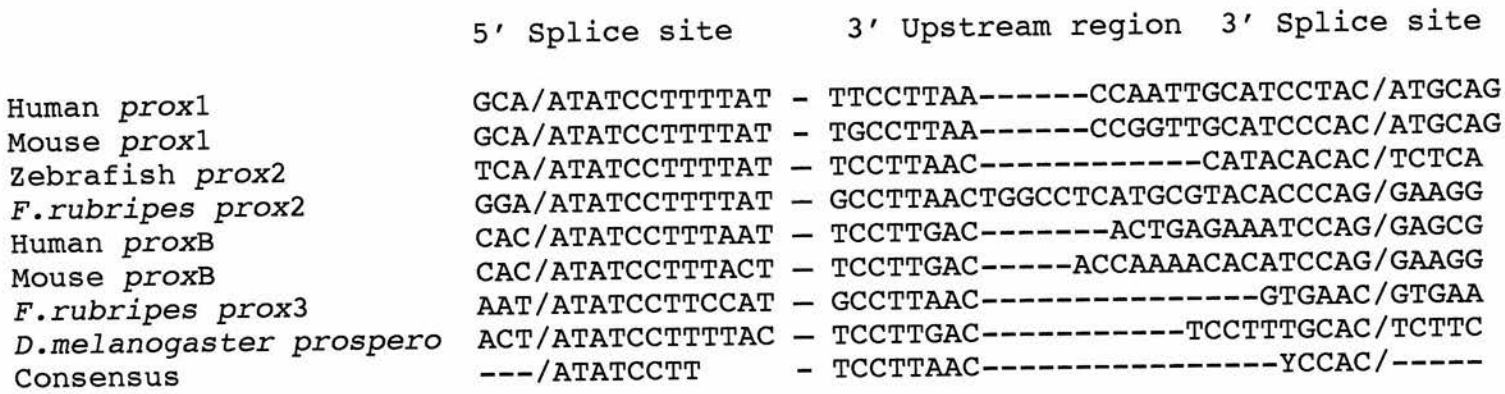



**Fig. 2.7 Conservation of U12 splice site in the Prox gene family**

The splice sites of the second intron were examined in the new members of the Prox family and compared to those in *prospero* and the known Prox genes. The consensus sequence was identified by Hall and Padgett (1994).

| | | | |
|---|---|---|---|
| Intronic open reading frame | AGS/ISFYQNLRDLFVSNEPCYKRFTRAEF | L | DIISHQTVLNHTH/SQE |
| | | INTRON | |
| Intron sequence | ggt/atttcaccgttgtagcttactacggt | c | gaaaccagcacac/tcg |
| | cgc/tctaaatgatttcaacgaagtcgcat | t | attgaacttaaca/caa |
| | ata/acttacgatgttatactcattttaat | t | caactaactctac/tgg |

**Fig. 2.8 Possible twintron structure of zebrafish *prox2***

Genewise was used to examine the splicing of the possible open reading frame sequence in intron two of zebrafish *prox2*. The U12 splice site is shown by a /. The longest open reading frame that has splice sites is shown with the amino acid and codon sequence split by the intron given between the 5' and 3' sequences. Other possible 5' and 3' splice sites were found using StrataSplice, and are shaded blue. The potential splice sites are in the correct phase to function as a pair.

### 2.2.9 Phylogenetic Analysis of the Prox Protein Family

A phylogenetic bootstrap consensus tree of the Prox protein family was created using the minimum evolution method (Fig. 2.9) The tree was calculated from a ClustalW alignment using Mega2.

All of the clades have significant bootstrap values. The Prox1 proteins are all positioned as would be expected from the evolutionary relationship of the four species. Without positive selection the fish duplicates would be expected to cluster together in the phylogeny, this does not occur, suggesting there has been selection for the divergence of the duplicated fish proteins. Furthermore although zebrafish Prox1 clusters with the tetrapod Prox1 proteins, none of the fish proteins cluster with the ProxB proteins. If duplication of the tetrapod *prox1* and *proxB* genes has occurred in the Actinopterygian lineage, both copies of the *proxB* and one of the *prox1* duplicates have diverged rapidly from their ancestral sequence.

### 2.2.10 Synteny

The genomic regions of the predicted and identified genes were examined for synteny. For the mouse and human sequences, the flanking genes were located in the Ensembl genome viewer, for *Fugu* the genomic sequence was analysed using NIX. *Fugu* sequence was used instead of zebrafish sequence as larger genomic contigs were available and the compaction of the *Fugu* genome reduces the distance between genes.

Analysis of the genomic sequence surrounding *Fugu prox4* identified the same flanking genes as Human and mouse *proxB* (Fig. 2.10, only human is shown as gene order is conserved between mouse and human), dihydrolipoamide succinyltransferase (DLST) and ZAP3. Less genomic sequence was available for *Fugu prox3*, but a second *Fugu* gene matching DLST was found to flank *Fugu prox3*

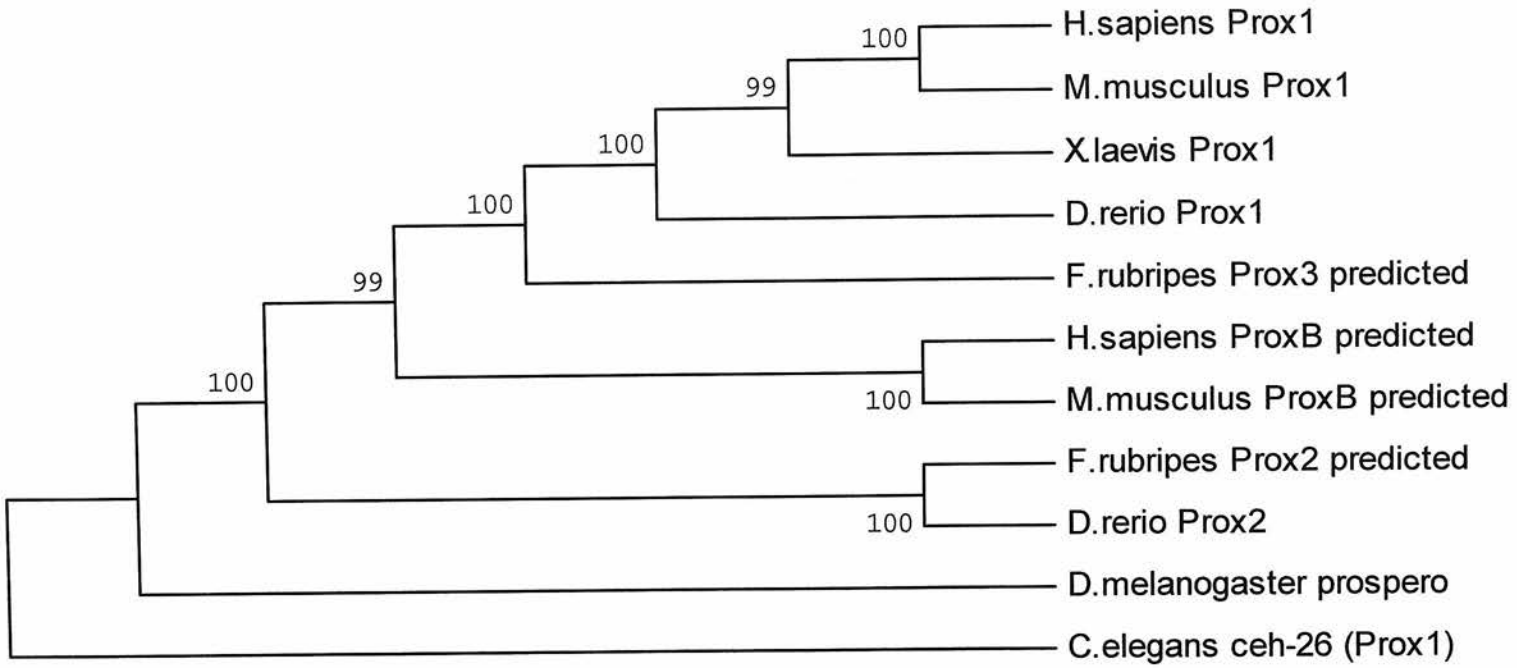



**Fig 2.9 Bootstrap consensus tree of the Prox protein family**

The values are percentage bootstrap values from 1000 repetitions, all of the values are significant. The Prox1 proteins all cluster as would be expected from the phylogeny of these species. The fish specific duplicates would be expected to cluster together, as they do not this suggests there has been selection for the divergence of these proteins. Also none of the fish proteins lie in the same clade as the mammalian ProxB proteins. This suggests that both of the fish duplicates of this protein have diverged rapidly since duplication.

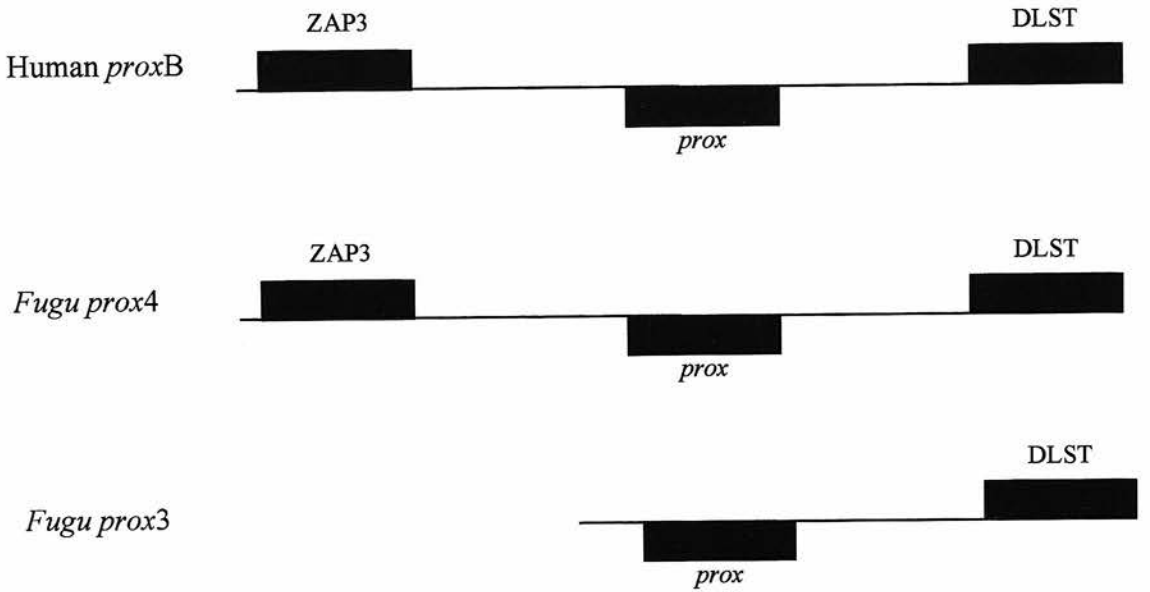



**Fig. 2.10 Synteny of *prox3* and *prox4* to *proxB***

NIX analysis of the genomic sequences identified dihydrolipoamide succinyltransferase (DLST) as a flanking gene to both *Fugu prox3* and *Fugu prox4*, and ZAP3 neighbouring *prox4*. Using the Ensembl genome browser ZAP3 and DLST were identified as the flanking genes to human *proxB* and mouse *proxB* (not shown).

on the opposite strand. This suggests that both *prox3* and *prox4* are syntenic to *proxB*.

Flanking genes were identified for *Fugu prox2* but no synteny could be found to either *prox1* or *proxB*.

## 2.3 Discussion

The results clearly demonstrate that the newly identified Prox proteins and Prox1 form a multiprotein family of *prospero* related transcription factors.

In order for duplicated genes to remain in the genome, rather than be inactivated through random mutation one of two events must occur. One of the genes may neofunctionalise, so that the duplicated genes now fulfil different functional roles. Alternatively they may subfunctionalise, splitting existing functions between the duplicated genes (Mazet and Shimeld, 2002). I also extend these descriptions to an embryonic rather than cellular level, such that subfunctionalisation may also include the splitting of expression between duplicated genes. In the case of the Prox genes, both situations may have occurred to some degree. In the zebrafish, there is both unique and overlapping expression for each of the *prox* genes. This demonstrates that there has been subfunctionalisation of the *prox* genes in zebrafish. However, all of the *prox* genes are expressed at their highest levels in the lens. This may indicate neofunctionalisation in the lens, or that the lens element of the promoter is essential for another function, such as expression in other tissues.

In both of the mammalian species examined, a second *prox* gene could be identified. The genes on mouse chromosome 12 and human 14q24.3 are syntenic to each other (Human-Mouse homology maps, see section 5.7.2). These genes also have the same genomic organisation, and cluster together in the phylogenetic analysis. It is clear that these genes are orthologous. The potential fourth family member, *Fugu prox4*, is

at the end of a genomic region demonstrated to be homologous to human 14q24.3 (Trower et al., 1996), and we also show that the synteny continues past *prox*4 to ZAP3. However, it appears from the synteny analysis that *Fugu prox3* may also be syntenic to human and mouse *proxB*. This is further supported by the evidence that *Fugu prox*3 has the same genomic organisation as the *prox*B genes. The *Fugu prox3* and *prox4* genes are, potentially, the result of a duplication of the *proxB* gene in the Actinopterygian lineage, as a result of which *prox*4 rapidly diverged from the other *prox* genes.

*prox4* could be identified in both *Fugu rubripes* and *Tetraodon nigroviridus* but not in zebrafish. It is possible that as more of the zebrafish genome is sequenced we will find evidence of this gene, but given the rapid divergence of this gene, it is also possible that this gene was lost in the zebrafish lineage but retained in the *Euteleostei* lineage.

Our evidence suggests that *prox3* and *prox4* are the result of a duplication of the *prox*B gene. The simplest model which explains the presence of two mammalian *prox* genes and four fish genes, given the two duplications in the vertebrate lineage and the duplication in the Actinopterygian lineage, is the loss of one copy following the first vertebrate duplication, and the retention of all copies from the remaining duplication events.

The U12 intron described in *prospero* appears to be conserved in all of the *prox* genes. Currently *prospero* is the only U12 containing gene for which the vertebrate homologues have been described to retain the U12 intron (Tomarev et al., 1998). This may be due to the presence of a twintron at the splice site, the conservation of amino acid sequence for the longer isoform preventing mutation of the U12 branch site. In all of the genes examined the U12 intron was phase 0 which would be required for a twintron to be present. In zebrafish *prox2* an open reading frame and possible U2 splice sites were identified nested within the U12 splice sites. We could not however obtain any evidence for alternative transcripts, therefore, whilst

genomic analysis identifies potential open reading frames and splice sites, it remains unclear if the twintron structure is maintained in the vertebrate homologues.

The Prox gene family is described here for the first time. Two *prox* genes were identified in the sequenced mammalian genomes. Given the duplication of the genome in the Actinopterygian lineage we would expect there to have been at most four genes in the fish species examined. We show evidence for the existence of all four genes in the *Fugu* genome. Therefore we believe that we have identified examples of all the members of the Prox protein family, including three members in zebrafish

### 2.3.1 Future Work

The identification of four Prox genes in teleosts and two in mammals provides a system in which the effects of duplication can be studied. The effects of duplication have been well studied in the Hox cluster (Amores et al., 1998; Prince et al., 1998; Wada et al., 1999; Manzanares et al., 2000; Bruce et al., 2001). The Hox cluster may not however, be representative of the effect of genome duplication on most genes, given that it is already a tandem array of duplications that share promoter elements. The expression of the Prox genes in zebrafish demonstrates subfunctionalisation. It will be interesting therefore to examine the genomic sequence surrounding the Prox genes and to try and identify promoter elements. Together with functional testing of any conserved regions this may give a better understanding of how promoters evolve as a consequence of duplication.

I have isolated a full-length cDNA for *prox2* and a partial sequence for *prox3*. In order to further study these genes I will experimentally verify the predicted sequence by isolating a full-length sequence of *prox3*.

The alignment of the Prox proteins demonstrates that the N-terminal region has very little constraint with the exception of a few motifs. By examining the ratio of

synonymous versus non-synonymous changes in this sequence it would be possible to determine if there had been selective pressure for amino acid changes within this region. If there were a significant increase in non-synonymous changes it would be evidence for neofunctionalisation.

My analysis of *prox2* suggested that a twintron could still be present, although no alternative transcripts have been described for any of the *prox* genes. It would be relatively simple to examine the possibility of a twintron splice structure using RT-PCR to detect the different isoforms.

The phenotype of the mouse Prox1 knockout has been described for the eye (Wigle et al., 1999), the lymphatic system (Wigle and Oliver, 1999), and the liver (Sosa-Pineda et al., 2000). Of the zebrafish genes identified, all three genes are expressed in the lens and both *prox1* and *prox2* are present in the liver. It is therefore possible that there could be redundancy between mouse Prox1 and ProxB, the phenotype in the lens and liver representing only a partial loss of function. Once we have isolated the 5' region of zebrafish *prox3* I will be able to knockout all Prox function using a triple morpholino approach. This would allow me to study the phenotype of a total loss of Prox protein.

# Chapter 3: Prox1 Function

## 3. 1 Introduction

The characterisation of the zebrafish Prox genes identified the embryonic slow muscle as a unique site of expression for *prox1*. Therefore, we decided to investigate the function of the Prox family members by studying the role of Prox1 in slow muscle formation.

Given that Prox1 is a developmentally expressed transcription factor we devised four hypotheses for its function in slow muscle:

- Prox1 may function to specify slow muscle fate.
- Prox1 may prevent the fusion of the slow muscle cells.
- Prox1 may play a role in the migration of the adaxial cells to form a superficial layer.
- Prox1 may be required for some aspect of the slow muscle structure.

As a prelude to investigating these hypotheses I further examined the expression of *prox1* in the slow muscle by in situ hybridisation. I then investigated its function using a combination of approaches. To determine if *prox1* functions to prevent the fusion of the slow muscle we overexpressed *prox1* specifically in the fast muscle using the fast muscle specific promoter of zebrafish fast myosin light chain 2 (*mlc2f*; Xu et al., 1999). We also used the new technique of morpholino knockdown to investigate the consequence of a loss of Prox1.

Slow muscle fate in zebrafish is usually determined by antibody labelling, in particular F59. The F59 antibody was raised against chick myosin heavy chains and demonstrated to label fast muscle in chick (Miller and Stockdale, 1986). Subsequently it has been demonstrated to label slow muscle early in zebrafish

development followed by, weaker, fast muscle staining (Devoto et al., 1996; Du et al., 1997). Other muscle antibodies have been demonstrated to have slow muscle expression, but again were selected for their labelling of myosin heavy chains rather than raised against specific epitopes. Therefore, we decided to isolate a definitive molecular marker of slow muscle differentiation for our assay of slow muscle fate. We chose slow myosin heavy chain (sMyHC) as a definitive marker of terminal differentiation in slow muscle and describe the identification and chararcterisation of a fragment of zebrafish sMyHC.

## 3.2 Results

### 3.2.1 *prox1* Expression

The Prox1 antibody was raised against the human Prox1 C-terminal region containing the homeodomain and *prospero* domains (Belecky-Adams et al., 1997), and has been shown to cross-react with chick, mouse, and zebrafish Prox1 (Belecky-Adams et al., 1997; Glasgow and Tomarev, 1998; Wigle et al., 1999). We have described the identification of multiple Prox proteins in zebrafish, all of which contain the highly conserved C-terminal region against which the antibody was raised. The in-situ patterns for *prox2* and *prox3* however demonstrate that the antibody is only recognising Prox1. We wished to examine the timing of expression in the muscle and therefore examined the expression of *prox1* by in situ hybridisation, using a probe to the full-length cDNA we had previously isolated (see section 2.2.1). At the 16-somite stage expression was clearly visible in the adaxial cells (Fig. 3.1a). At 24 hours (26-somites) expression was in the adaxial cells posteriorly, and in migrating and lateral slow muscle in more anterior somites (Fig. 3.1b). Sections through the embryo demonstrated that staining was restricted to the single layer of lateral slow muscle cells after migration (Fig. 3.1c).

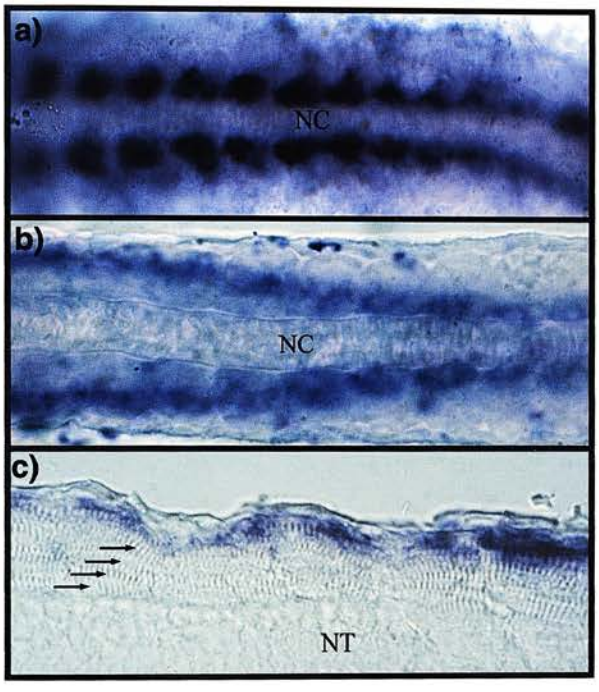



**Fig. 3.1 *prox1* muscle expression**

In situ hybridisation with a *prox1* probe.
a) A 16-somite embryo showing prox1 expression in the adaxial cells either side of the notochord (NC).
b) Posterior somites in a 26-somite embryo showing expression in the lateral slow muscle, the migrating slow, muscle and the adaxial cells.
c) A section at the level of the neural tube (NT) showing the single layer of *prox1* expressing slow muscle. The fibres of the fast muscle (arrows) are clearly seen and do not express *prox1*. Expression is decreasing in the anterior, older, somites (the section was cut by K. Steveling).

### 3.2.2 Identification and Characterisation of a Zebrafish Slow Myosin Heavy Chain

In order to assay slow muscle fate we isolated a definitive marker of slow muscle terminal differentiation, slow myosin heavy chain sMyHC. We searched the EST database with a partial fragment of trout slow myosin heavy chain (AAF74412) and identified multiple ESTs with homology to the trout protein. Two of these ESTs were then partially sequenced by D. Keenan. The sequence reads from these ESTs and sequences from ESTs in the same cluster, as identified by the WashU-Zebrafish Genome resources, were assembled into three contigs (Appendix 1.2.3). To further clarify the type of myosin we had isolated, we carried out phylogenetic analysis of fish myosin heavy chains (Fig. 3.2). Sequences were included in the analysis only if they had been previously characterised by expression or contractility, not just on the basis of sequence similarity (carp 30°C isoform BAA22069, carp 20°C isoform BAA22068, carp 10°C isoform BAA22067, white croaker BAB12571, Alaskan pollack BAA19070, freckled hawkfish CAC59753, zebrafish AAK73348, and *Drosophila* P05661).

The regions of the sequence corresponding to the fragment available for the trout protein were used to construct a phylogenetic tree. The tree was constructed using the minimum evolution method with complete deletion, and bootstrapped using a random seed and 1,000 repetitions. The tree demonstrates that the zebrafish myosin heavy chain we identified clusters with the trout slow myosin. This suggests that the fragment we isolated is from a zebrafish sMyHC. To confirm that the fragment was a sMyHC we characterised the expression of the zebrafish EST isolated. An antisense in situ probe was synthesised using the EST as a template (the in situ probe was made by D.Daggett). The EST was expressed in the adaxial cells at the 8-somite stage (Fig. 3.3a), adaxial and migrating muscle at the 17-somite stage and at the 28-somite stage the superficial layer of slow muscle anteriorly (Fig. 3.3c-d) with adaxial and migrating. Given the phylogenetic analysis and expression pattern we believe this to be a fragment of zebrafish sMyHC. The expression pattern confirms that the probe could be used as a marker of terminally differentiated slow muscle.

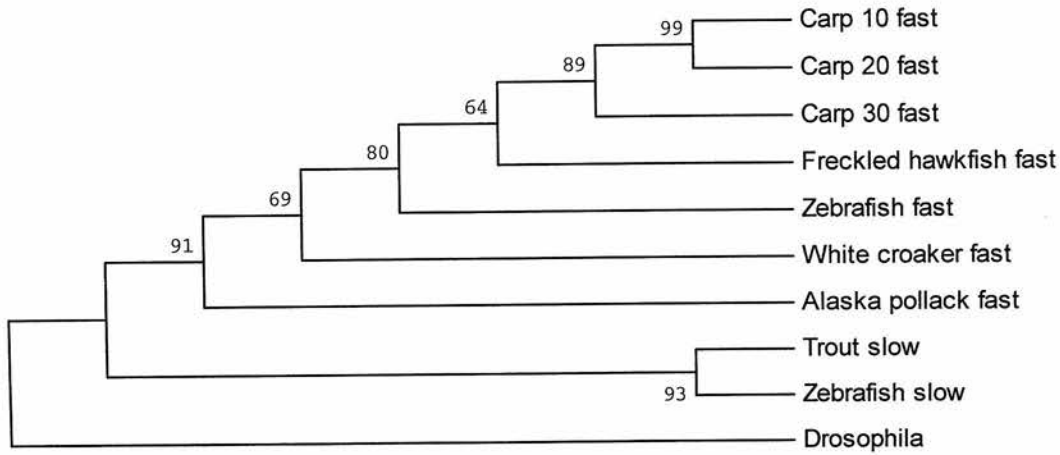



**Fig. 3.2 Phylogenetic tree of fish myosin heavy chains**

A phylogenetic tree was created from the alignment of myosin heavy chains regions to the fragment available from trout. *Drosophila* myosin heavy chain was used to root the tree. The values given are percentage bootstrap values from 1000 repetitions, and the tree was created using the minimum evolution method. The trout slow myosin and zebrafish myosin form a separate clade to all of the identified fast myosin heavy chain proteins. This suggests that the zebrafish protein is also a slow specific isoform.

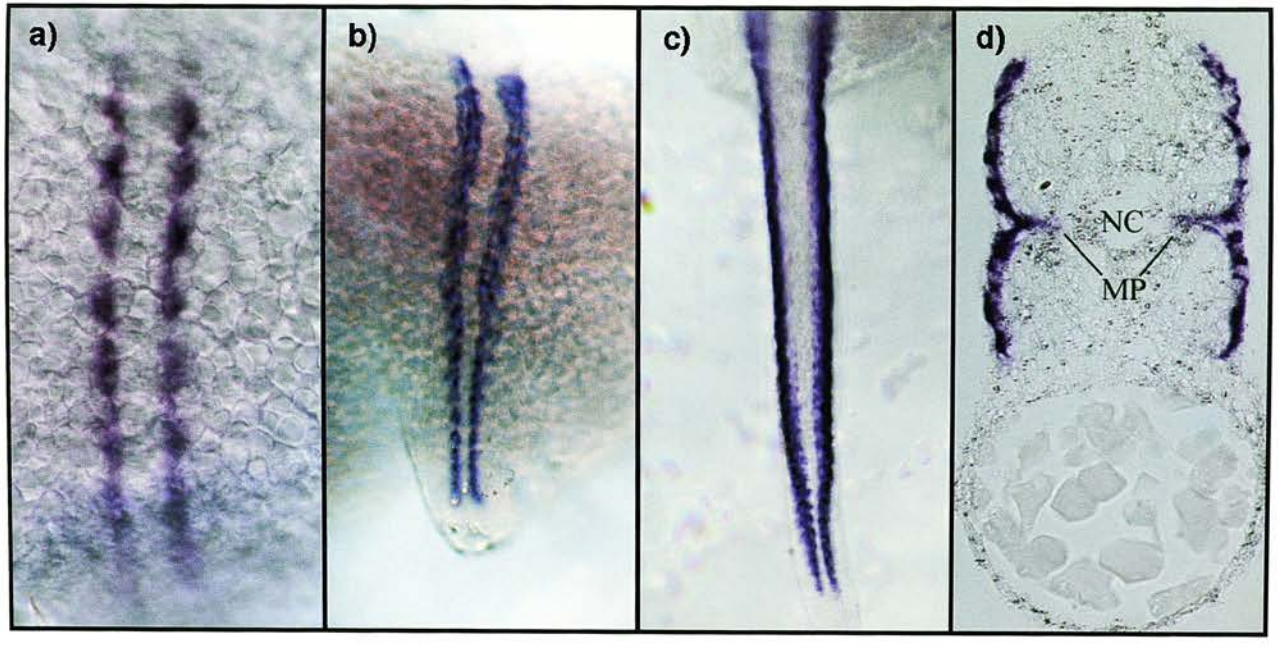



**Fig. 3.3 Expression of slow myosin heavy chain**

a) At the 8-somite stage expression is clearly visible in the adaxial cells either side of the notochord.
b) Expression at the 17-somite stage in the adaxial cells prior to migration.
c) At the 28-somite stage expression is seen anteriorly in the lateral post-migratory slow muscle. Posteriorly expression is visible in the migrating slow muscle, and in the most posterior somites in the adaxial cells.
d) Transverse section through a 28-somite stage embryo at the level of the yolk extension, showing expression in the lateral slow muscle and more medially in the muscle pioneer cells (MP; NC, notochord).

### 3.2.3 Knockdown of Prox1

To examine the consequence of a loss of Prox1 function we inhibited *prox*1 translation using a morpholino oligonucleotide (reviewed in Summerton and Weller, 1997). Morpholinos are 25 base pair antisense oligonucleotides complementary to the sequence of the 5' untranslated region of a gene up to the ATG. The morpholino inhibits translation initiation, preventing synthesis of the targeted protein. The morpholinos have morpholine rings, joined by non-ionic phosphorodiamidate linkages, in place of ribose or deoxribose present in the backbone of RNA or DNA. One consequence of this is that the oligonucleotide becomes resistant to nucleases, making it very stable (Hudziak et al., 1996). Morpholino injections in zebrafish have been shown to prevent target protein expression and reproduce null phenotypes of known mutants (Nasevicius and Ekker, 2000).

To determine if Prox1 function was required for the medial-lateral migration of the slow muscle we examined the position of the slow fibres in embryos injected with a morpholino oligo against *prox1*. Injected embryos were examined for a phenotype at 24 hours (n=53); 45% of the embryos demonstrated a blocky somite phenotype. The phenotypic embryos (n=24), together with wild type age matched embryos and non-phenotypic injected siblings, were labelled with the F59 antibody, to determine the position of the slow muscle fibres, and Prox1 antibodies, to demonstrate the loss of Prox1 protein.

The stained embryos were examined by confocal microscopy. Morpholino injected embryos show a complete loss of Prox1 expression (n=6) compared to wild type and non-phenotypic injected embryos (n=6), where the Prox1 positive nuclei were clearly visible in the centre of the slow cell (Fig. 3.4). The blocky somite phenotype was evident in the morphant embryos (Fig. 3.5a-c), having rounded somites and poorly defined somitic boundaries compared to the chevron shaped somites of the wild type (Fig. 3.5d). No difference was observed in the position of the F59 positive fibres between morpholino injected and wildtype embryos, indicating that the medial-lateral migration of the adaxial cells was unaffected. The patterning of the myofibres

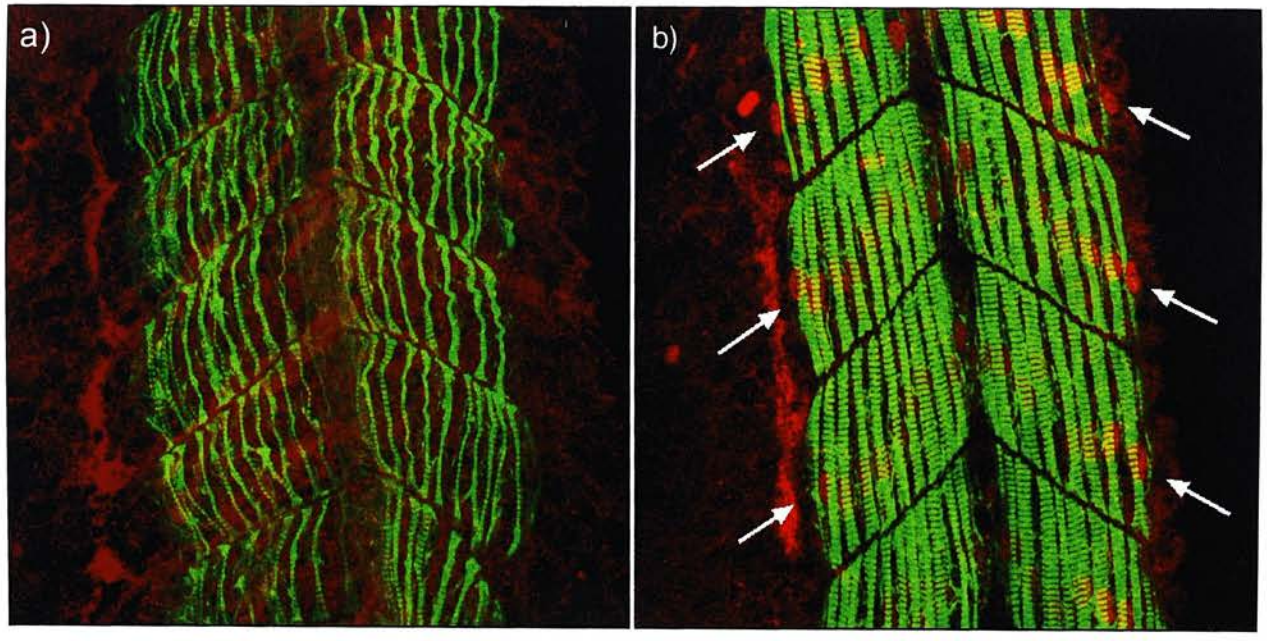



**Fig. 3.4 *prox1* morpholino injected embryo**

Double labelling of slow muscle using the F59 antibody (green) and the Prox1 antibody (red). Embryos are at the 24-somite stage. Embryos are shown in lateral views, anterior to the top. Images are a stack of confocal sections.
a) *prox1* morpholino injected embryo. No Prox1 staining is visible and the slow muscle staining is reduced. The somite boundaries are present but the somites fail to take on a chevron shape. The level of F59 staining is reduced compared to wildtype and the diameter of the fibres appears to be reduced. Sarcomeric banding is still apparent but the myofibres are disorganised.
b) wildtype embryo showing strong F59 staining of the slow fibres. The fibres are have an organised palisade-like appearence with very little space between the fibres. Prox1 staining is clearly visible in the centrally located nuclei (arrows).

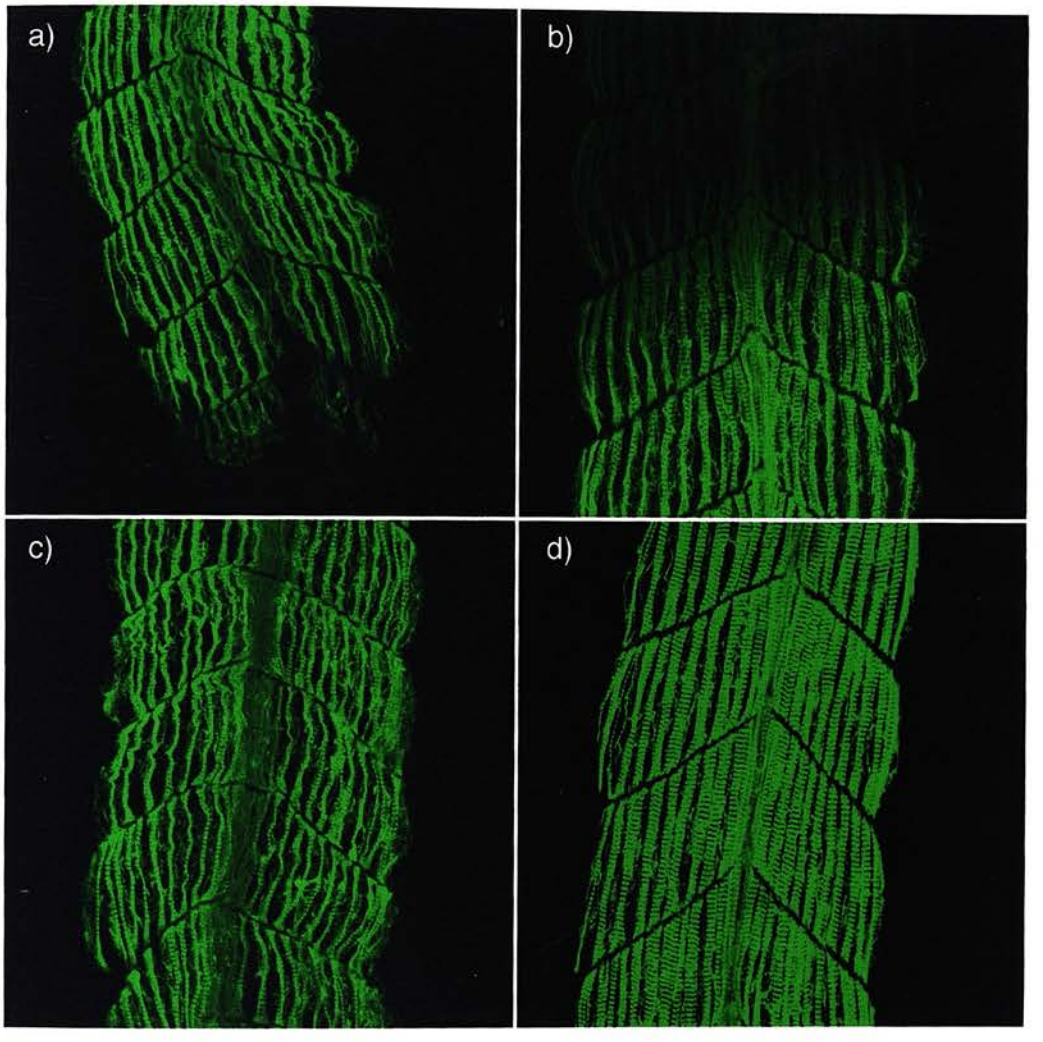



**Fig. 3.5 Slow muscle fibres in *prox1* morpholino injected embryos**

26-somite stage embryos, slow muscle fibres are labelled with F59 (green), lateral views, anterior to the top. The images are produced from stacks of confocal sections.
a-c) *prox1* morpholino injected embryos. The patterning of the slow fibres is disrupted, fibres are smaller in diameter and are wavy in appearance, and the somites do not fully form into a chevron shape.
d) wildtype embryo showing strong F59 staining. The fibres are regularly patterned with little space between fibres, and the chevron shape of the somite is clear.

did appear to be disrupted compared to wild type, and the level of staining reduced (Fig. 3.5). All of the fibres appeared to be attached at the somite boundaries and running anterior-posterior, but the fibres have a wavy appearance in contrast to the perfectly straight fibres of the wild-type embryo. Some of the fibres in the morpholino injected embryo can also be seen to split (Fig. 3.5a-c).

F59 labels slow cells at 24 hours, but later in development labels both slow and fast muscle. To determine if F59 expressing cells were still slow muscle, or if they had changed fate to become early differentiating fast muscle, we carried out in situ hybridisation with sMyHC. In all of the embryos examined (n=18) there was no change in the sMyHC staining pattern compared to wild type (Fig. 3.6).

### 3.2.4 Fast Muscle Expression of *prox*1

In addition to removing *prox1* function through morpholino injection, we also examined the consequence of ectopic *prox1* expression in the fast muscle. We mosaically expressed *prox1* in the fast domain using the *mlc2f* promoter (Xu et al., 1999). Skeletal myosin consists of two myosin heavy chain and four light chains, *mlc2f* has been characterised as a fast muscle specific regulatory light chain. The promoter of *mlc2f* has been demonstrated to drive fast muscle specific expression (Xu et al., 1999), and contains several E-box motifs, known binding sites for myogenic transcription factors, including a consensus MEF2 binding site, which was demonstrated to be sufficient to drive expression in the muscle (Xu et al., 1999).

The *prox*1 coding sequence was cloned into the *mlc2f*promoter-IRES-GFP vector (the *mlc2f*promoter-IRES-GFP vector was made by F.Cortes), to create a *mlc2f*promoter-*prox1*-IRES-GFP construct (Fig. 3.7). This construct would express both *prox1* and *GFP*, through the use of the internal ribosome entry site (IRES) sequences, in the fast muscle.

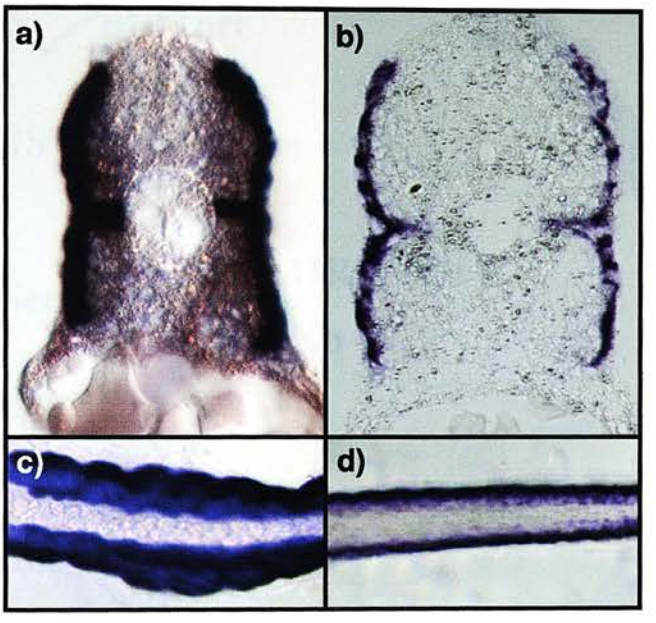



**Fig. 3.6 Slow myosin heavy chain staining of *prox1* morpholino injected embryos**

28-somite stage embryos labelled with slow myosin heavy chain.
a,b) Transverse section showing stong labelling of the superficial slow muscle and the adaxial cells immediately adjacent to the notochord in both a *prox1* morpholino injected (a) and wildtype embryo (b).
c,d) Dorsal view showing staining of the slow muscle in morpholino injected (c) and wildtype embryo (d).
Sections were cut by C. Neyt.

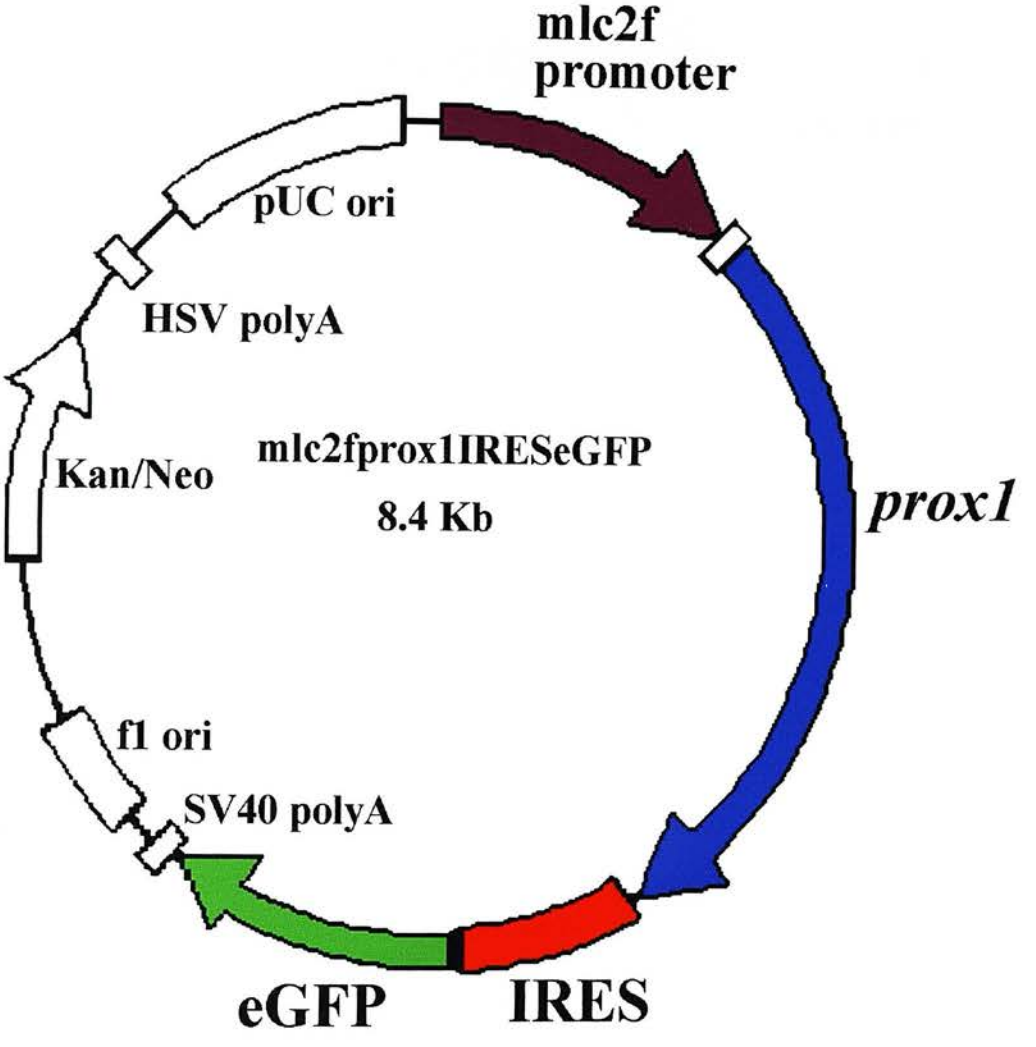



**Fig. 3.7 *mlc2f-prox1*-IRES-GFP construct**

Injection construct expressing *prox1* and GFP under the control of the *fast myosin light chain 2* (*mlc2f*) promoter. Both Prox1 and GFP are transcribed through the use of an internal ribosome entry site (IRES). The original *mlc2f*-IRES-GFP construct was made by F. Cortes.

Embryos were examined at 48 hours and the fusion status of the embryos determined by propidium iodide nuclear staining. GFP expressing embryos were analysed using confocal microscopy. GFP positive cells all contained multiple nuclei (n=9), GFP expressing cells and surrounding wildtype cells having on average four nuclei. However, the nuclei of GFP positive cells were centred in the syncytium, compared to the even distribution of nuclei in the surrounding fast muscle cells (Fig. 3.8a-c). This effect was never observed in GFP positive cells in control embryos injected with *mlc2f*promoter-IRES-GFP (n=15, Fig. 3.8d).

## 3.3 Discussion

In the morpholino experiments the slow muscle is still specified, as demonstrated by F59 and sMyHC staining. This demonstrates that Prox1 is not necessary for the specification of the slow muscle. Additionally the position of the slow fibres is unaffected by the loss of Prox1, demonstrating that the medial-lateral migration of the non-pioneer slow muscle cells is unaffected. The expression of Prox1 in the fast muscle does not prevent fusion occurring, suggesting that its function is not to preserve the mononucleate nature of the slow fibres. The fast muscle expression supports the results of the morpholino injections, as there is also no lateral migration of the Prox1 expressing fast cells to occupy a position characteristic of the slow cells. The results demonstrate that Prox1 is not necessary for the specification of a slow fate, as assayed by the expression of myosin heavy chains, and is neither necessary nor sufficient for the migration of the non-pioneer slow muscle cells.

Whilst Prox1 does not specify a slow fate it may be required for some aspect of slow muscle morphogenesis. An interesting observation is that the nuclei of the fast cells appear to cluster at the centre of the syncytium in Prox1 expressing cells. A central nuclear position is characteristic of the mononucleate slow cells, and it is possible that Prox1 plays a role in this localisation, although no functional significance of this localisation is known.

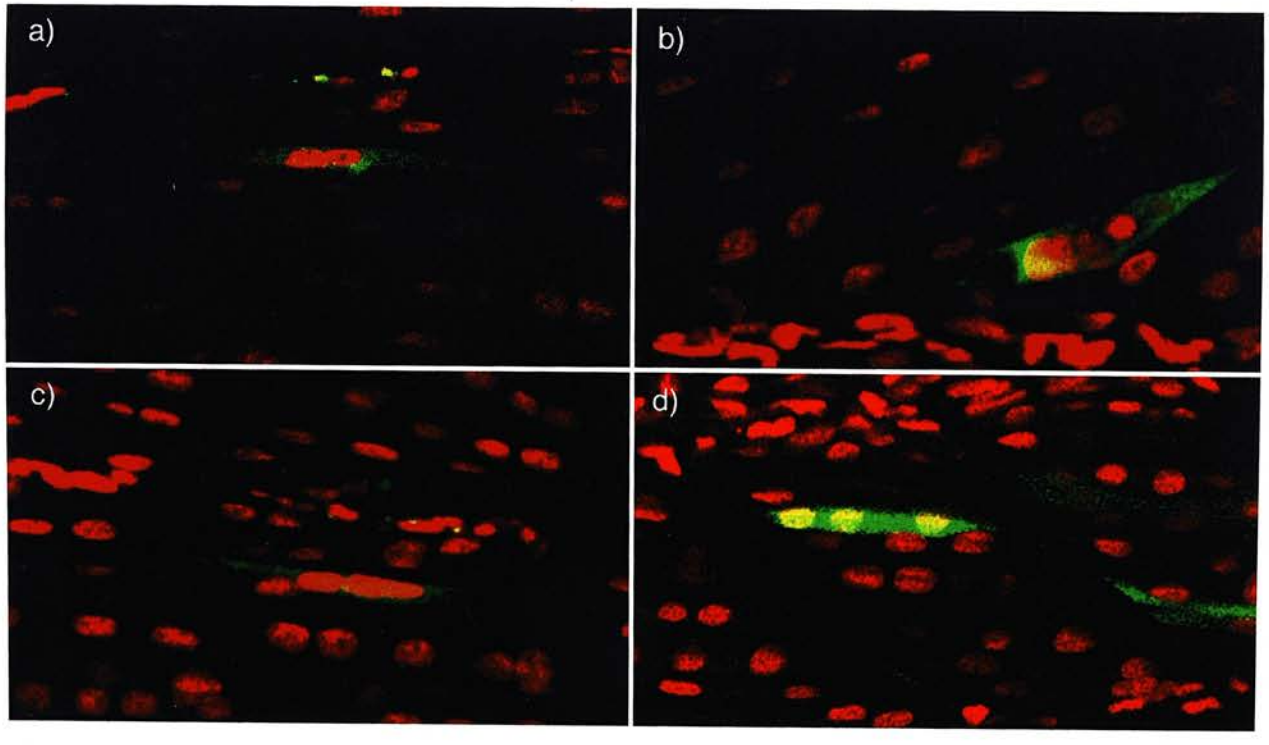



**Fig. 3.8 *prox1* and GFP ovexpression in the fast muscle**

Propidium iodide labelled embryos. Lateral views anterior to the left.
a-c) *mlc2f-prox1*-IRES-GFP injected embryos. *prox1* expressing cells, labelled by GFP expresison (green) have the same number of nulei (red) as surrounding fibres. However the nuclei in *prox1* expressing cells are centrally located in the syncytium.
d) *mlc2f*-IRES-GFP injected control embryo demonstrating the wildtype even spacing of nuclei.

The disruption of muscle fibres in the morpholino injected embryos suggests a role for Prox1 in the patterning of the slow muscle fibres. The nuclei phenotype in the fast cells may therefore be a consequence of altered patterning or structure, rather than a specific role in controlling nuclear position.

During the course of this work Roy and colleagues (2001) described the loss of Prox1 in the *u-boot* (*ubo*) mutant. In order to investigate the role of Prox1 in the *ubo* mutant Prox1 was removed by morpholino knockdown. The results concur with what I have discovered, no effect is seen on the expression of MHC as assayed by F59 but myofibrillar organisation is disrupted (Roy et al., 2001). Roy and colleagues (2001) report the random orientation of slow fibres in the morpholino injected embryos. I do not observe this random orientation in my data; all fibres are orientated in an anterior-posterior direction with the sarcomeres perpendicular to the fibre. The fibres are no longer perfectly straight, instead having a wavy appearance, and in some cases the fibres appear to split.

In conclusion my data suggests that Prox1 has a role in the patterning of the slow fibres and may be required for the initiation of a late1 step in terminal differentiation involving the patterning of the slow fibres.

### 3.3.1 Future Work

The results of the morpholino injections demonstrate a role for Prox1 in the correct patterning of the slow fibres. The Sev signalling pathway, which has been demonstrated to regulate *prospero* in the *Drosophila* eye through the Ras/Raf/MAPK signalling pathway (Kauffmann et al., 1996; Xu et al., 2000), has been implicated in cytoskeletal patterning. The mammalian homologue of *Drosophila* son-of-sevenless (mSos), a Ras/Rac guanine nucleotide exchange factor that mediates Sev activation of the Ras/Raf/MAPK signalling pathway (Karlovich et al., 1995), has been demonstrated to interact with regulators of the actin cytoskeleton (Wasiak et al., 2001). As well as signalling through mSos regulating the cytoskeleton, cytoskeletal

rearrangement has been shown to cause signalling through Sos/Ras/Raf/MAPK (Irigoyen et al., 1997). Pros is known to be regulated by signalling through the Sos/Ras/Raf/MAPK pathway and may contribute to the effect of this pathway on the cytoskeleton. In order to investigate the role of Prox1 in cytoskeletal patterning, I will examine the pattern of actin-filaments in wildtype and morphant embryos using phalloidin staining.

Examination of the Prox1 knockout mouse identified higher proliferation in the lens and down regulation of the cell cycle inhibitor $p27^{Kip1}$ (Wigle et al., 1999). Additionally, in *Drosophila, pros* mutants exhibit an increase in mitotic activity (Li and Vaessin, 2000). These suggest that Prox1 may also have a role in the regulation of the cell cycle. Cell cycle stage specific antibodies for M-phase and $G_1$/S phase have been demonstrated to function in zebrafish (Link et al., 2001), and I will use these to examine the cell cycle status in Prox1 morpholino injected embryos.

Work in culture has shown than disruption of the cytoskeleton can control regulation of the cell cycle (Huang et al., 1998). In particular cells that treated using cytoskeletal inhibitors to disrupt the actin filaments down-regulate the cell-cycle inhibitor $p27^{Kip1}$ (Huang et al., 1998). It is possible that the requirement of Prox1 in the structure of the slow fibres is through the patterning of the cytoskeleton, and that this patterning could also be linked to a role of Prox1 in cell cycle regulation.

The crystallin genes are good candidates for possible targets of Prox1, given that γ-crystallin has already been identified as a target (Lengler et al., 2001). The expression of αB-crystallin in the muscle has been well characterised in the mouse (Benjamin et al., 1997). The regulation of αB-crystallin has been well studied, and enhancer regions required for skeletal muscle expression identified (Dubin et al., 1991; Gopal-Srivastava and Piatigorsky, 1993). Expression of MyoD and Myogenin in non-muscle cell cultures have been demonstrated to upregulate αB-crystallin through an myogenic regulatory factor (MRF) binding site, although transcription was reduced compared to transfection into muscle cells in culture suggesting that other factors were also involved (Gopal-Srivastava and Piatigorsky, 1993). It is

possible that Prox1 may regulate αB-crystallin as well as γ-crystallin, therefore future work will include identifying crystallin genes and their expression. Preliminary work has identified α, β, and γ-crystallins in the zebrafish EST database, and the expression of an αA-crystallin has already been reported (Runkle et al., 2002).

### 3.3.2 Conclusions

We demonstrate that the overexpression of *prox1* leads to a disruption of slow muscle formation. Using a loss of function approach and the targeted expression of *prox1* in the fast muscle we rule out the requirement of Prox1 in the specification of slow fate, the migration of the non-pioneer slow cells, and in the prevention of slow muscle cell fusion. Furthermore we describe the isolation of a fragment of zebrafish slow myosin heavy chain and demonstrate its use as a definitive marker of slow muscle fate. We demonstrate that ectopic *prox1* expression disrupts the normal patterning of the fast muscle syncytiums and identify a role for Prox1 in the correct patterning of the slow fibres.

# Chapter 4: The Development of Optical Projection Tomography for use in Zebrafish

## 4.1 Introduction

Information about the spatio-temporal patterning of an embryonic structure, gene, or protein expression pattern can be invaluable in elucidating function. The determination of the 3-dimensional (3D) organisation of an anatomical structure within an embryo has typically required generating a reconstruction from serial sections (Brune et al., 1999; Streicher et al., 2000; Weninger and Mohun, 2002). The process of sectioning can damage the morphology of an embryo and cause the loss of 3D organisation. Despite the optical clarity of zebrafish embryos, sectioning is still required for the characterisation of gene and protein expression patterns. Some of these problems can be overcome by optical sectioning, such as confocal microscopy, however there are limits to the depth and size of tissues that can be examined preventing analysis after a certain stage of development. Mounting embryos for these procedures can also distort morphology. Furthermore, confocal analysis is only possible on fluorescent stains.

In order to overcome these problems we have extended the Optical Projection Tomography technique described by Sharpe *et al.* (2002) to allow its use on specimens as small as zebrafish embryos. Instead of taking 2-dimensional (2D) optical sections through the specimen, OPT allows 3D reconstruction from a series of images captured at multiple angles. Once reconstructed the embryo can be virtually sectioned in any plane and the morphology or expression pattern determined. Alternatively, the data can be surface or volume rendered, as described later, to produce 3D objects. These objects can then be viewed from any angle, allowing the visualisation of the sample in 3D.

### 4.1.1 The Theory of Projection Tomography

Projection tomography is the process by which information about an object is determined from the object's projections. Projection tomography has been used successfully for x-ray scans (Hounsfield, 1973), ultrasound, and magnetic resonance imaging (Lauterbur, 1973). In all of these methods a 3D object is produced from reconstructions using 2D projection images.

Examining the projection of an object removes a dimension as demonstrated in Fig. 4.1. For the 2D object in Fig. 4.1 the projection has a single dimension. We could imagine the same happening to a 3D object such that a 2D projection would be obtained. From a single projection image we can not reconstruct the 3D shape of the object, but if we obtain projection images from a series of angles it is possible to calculate the shape of the object as demonstrated in Fig. 4.2.

The object is not reconstructed as a 3D object but as a series of 2D sections. The sections are subsequently stacked to form a 3D object. The reconstruction process used is a back projection method. In the simplest form of back projection, the projection images are projected back through the object and the combination of the images reconstructs the shape of the original image, as shown in Fig. 4.2e). A consequence of this is that a star-shaped pattern forms around each object.

The Radon transform (Radon, 1917) is a projective transformation that generates a projection from a 2D object. Using the inverse of the Radon transform we can generate 2D information from the projection. This method has none of the artefacts created when using a simple back projection.

The result of the inverse Radon transformation is still blurred compared to the original object and so filtering is carried out to reduce this effect. A ramp filter is applied to the data either before or after back projection. The ramp filter attenuates data according to its frequency, with greatest effect on low frequency

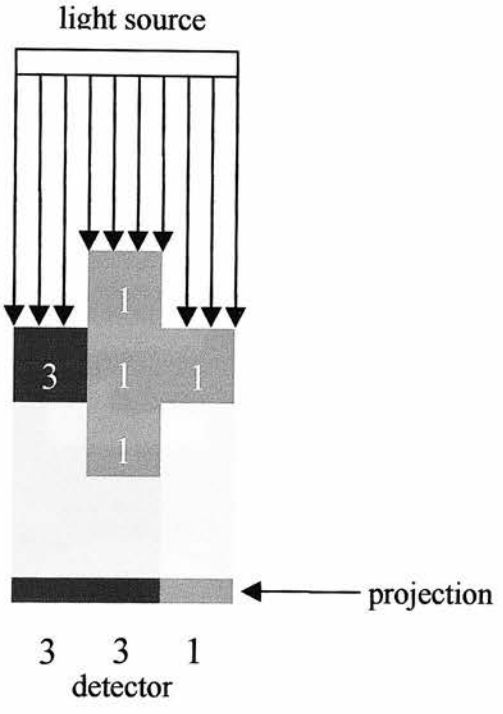



**Fig. 4.1 Generating Projection Images**

Light from the source passes through the object. Where the object is thicker less light passes through the object and the projection recorded on the detector is darker. Also where the object is denser, as shown by the density values on the diagram, less light reaches the detector. The value in the projection is the sum of density values along the light path through the object. A dimension of the sample is removed in the projection resulting in a single dimensional projection from a 2D sample.

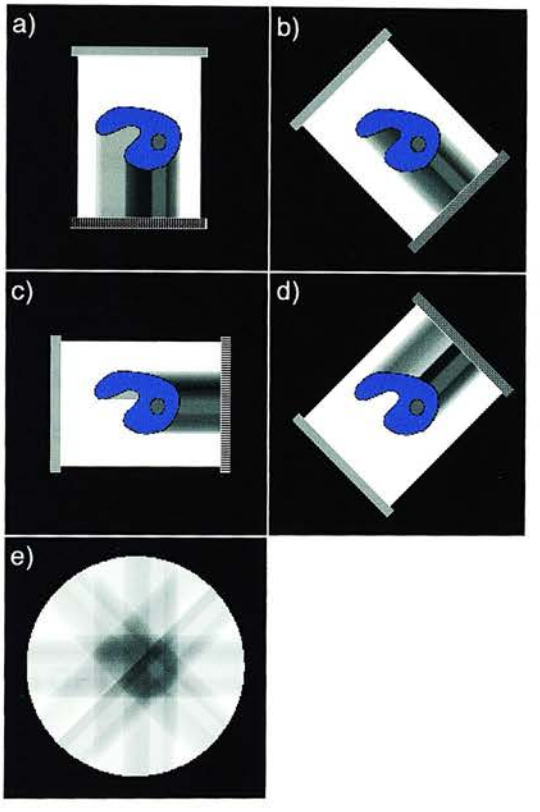



**Fig 4.2 Projection images and simple back projection**

a)-d) Show the capture of projection images at four angles around the same object. Where the object is thicker the projection is darker.
e) In the simplest form of reconstruction the projection images are projected back along the angles they were collected from. A star-shape pattern is clearly visible around the recontruction.
This figure is adapted from an image by J. Sharpe.

data and no attenuation of the highest frequency data. The frequency refers to the rate of change of intensities in the projection (i.e. a very bright pixel next to a black pixel has a very high frequency, whilst a line of pixels with very similar values has a low frequency). This enhances edges in the reconstruction, making the results sharper. In most cases, including OPT, the filter is applied to the projection before reconstruction, as the application of a 1-dimensional filter is less computationally intensive than a 2D filter.

Using the projections gathered over a series of angles results in an object that has been reconstructed along radial lines (Fig. 4.3). The data is interpolated to give a reconstruction on a square grid, at the outside of the circle the data points are further apart resulting in larger interpolation errors. Therefore the reconstruction is most accurate at the axis of rotation. In order to reconstruct the object from the series of projections, information must be obtained over at least 180°. As the number of angles used increases, the quality of the reconstruction improves as the distance between data points is reduced.

### 4.1.2 Zebrafish OPT

The smallest sample that had been successfully analysed by OPT, prior to these experiments, was a 9.5 dpc mouse embryo. These embryos measure approximately 2.5mm long and 1.6mm wide (Kaufman, 1992), compared to zebrafish embryos at 28-somites which are approximately 1.6mm long and only 160μm wide. Therefore a significant improvement in resolution was required to successfully reconstruct a zebrafish embryo.

In order to test the potential of OPT for use in zebrafish we chose to use an anti-acetylated tubulin antibody. Acetylated tubulin is one of the first components of the neuronal microtubule cytoskeleton to form and is a marker of developing post-mitotic neurons (Piperno and Fuller, 1985; Wilson and Easter, Jr., 1991). This

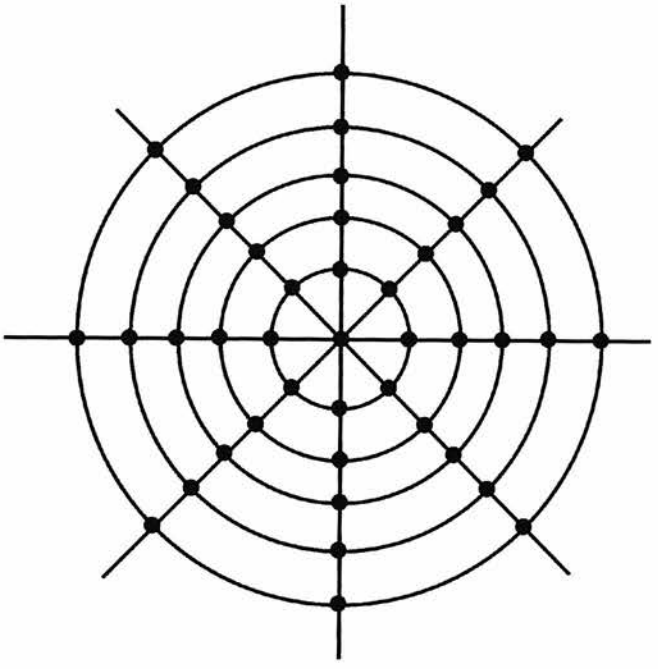


**Fig. 4.3 Reconstructions are generated radially**

The back projection of the projection images produces a radial reconstruction. The straight lines each represent the data from a single projection image. In total four projection images over 180° are shown, with 11 points of data in each projection. Interpolation of the data allows the reconstruction to be positioned on a square grid. At the axis of rotation the data points are close together reducing the interpolation error, at the outside of the object the data points are further apart and the interpolation error will be greater, reducing the quality of the reconstruction. Adapted from Kak and Slaney 1988.

reveals the very fine pattern of neuronal cells bodies and fine processes in the developing embryo, which would help to determine the possible resolution of OPT.

Images may be captured for OPT using brightfield or fluorescent imaging. The embryos are cleared to allow light to pass through the sample, so the projection images can be captured. Initial experiments showed that the optical clarity of the zebrafish embryo, when cleared, prevented brightfield imaging, as the outline of the embryo was hard to distinguish and internal details were lost. In the larger mouse embryos this was not a problem, as there was still a sufficient reduction of light passing through the embryo to distinguish embryonic tissue from the background. If anatomical features and the outline of the embryo are not distinguishable, the location of the staining pattern within the embryo can not be determined. An alternative is to use fluorescent imaging to capture the anatomical features of the embryo by their autofluorescence. Additionally in mice embryos fluorescent OPT produced sharper images than brightfield OPT (Sharpe et al., 2002). The weak autofluorescence of zebrafish embryonic tissue allowed the outline of the embryo to be captured, but very little internal detail. Despite the improvement over brightfield imaging, the long exposure times required due to the low level of autofluorescence resulted in poor quality images and the high level of yolk autofluorescence obscured many details.

We developed techniques to solve these problems and present their application in zebrafish, demonstrating their use as a novel, convenient and quick method for obtaining 3D data that could not be obtained using conventional techniques. We believe this could become a new standard for the presentation of anatomical and expression data.

## 4.2 Results

### 4.2.1 Sample Preparation

In zebrafish embryos the autofluorescence of the yolk is much greater than that of embryonic tissue in unstained embryos. This prevents accurate reconstructions using the standard protocols, as the yolk obscures embryonic structures. In antibody labelled embryos the staining procedure reduces the yolk fluorescence, but background embryonic fluorescence remains very weak. As a consequence embryonic structures can not be reconstructed and the localisation of staining within the embryo cannot be determined.

In order to overcome this problem we tested different fixation techniques and searched for fluorescent dyes to increase the fluorescence of the entire embryo. We tested 4% paraformaldehyde (PFA), 10% formalin, and 10% formalin containing 0.25% gluteraldehyde. Whilst gluteraldehyde increase the fluorescence of the embryo it had a stronger effect on the surface of the embryo and therefore PFA was chosen, as this produced a more uniform effect. Propidium iodide (PI) binds to nucleic acids and fluoresces strongly with a broad emission at the red end of the spectrum. PI staining increased the signal from the embryonic tissue above that of the yolk, and greatly reduced the exposure times required. Embryos were washed for 1 hour at room temperature in PI at a concentration of 0.2µg/ml. The embryos were then washed 5 times in PBT at room temperature.

28-somite stage zebrafish embryos were stained with an acetylated tubulin antibody (Piperno and Fuller, 1985) and an Alexa488 (Molecular Probes) coupled secondary antibody before staining with PI. The embryos were then mounted vertically in low melting point agarose at 30°C which was allowed to set. The agarose mount was dehydrated by immersion in methanol overnight and then cleared using 2:1 benzyl benzoate: benzyl alcohol for at least 5 hours.

### 4.2.2 Image Capture

Images were captured using a Leica MZ FLIII stereomicroscope with a photometrics coolsnapcf digital camera (Roper Scientific) and the 'Edinburgh OPT Scanner', which holds and rotates the specimen so it can be imaged.

For mice embryos, 12-bit images were captured at a resolution of 696 X 520 pixels. Due to the smaller size of the zebrafish embryos it was decided to capture images at the highest resolution available to the camera, 1392 X 1040 pixels. A series of 400 images were captured over a 360° rotation of the sample. In order to reduce the noise from the digital camera, background subtraction was applied to the images. An exposure of the same duration was captured but with no light being allowed to pass to the camera. The values in this background image were then subtracted from all of the images captured during the scan. This removed aberrations due to differences in sensitivity of individual pixels in the digital camera. Image acquisition and background subtraction was carried out using IPLab software (Scanalytics). Examples of images from the scan are shown in Fig. 4.4 and the supplementary data.

### 4.2.3 Reconstruction

The images were then put through the reconstruction pipeline, composed of a series of individual programs. The pipeline was designed to produce a reconstruction requiring the least interpolation of data, and assessed empirically. Programs were developed by James Sharpe (JS) or the Mouse Atlas Project (MA), authorship is given the first time that the program is referred to. Programs not labelled (JS) or (MA) were written by the author and are presented in the Appendix.

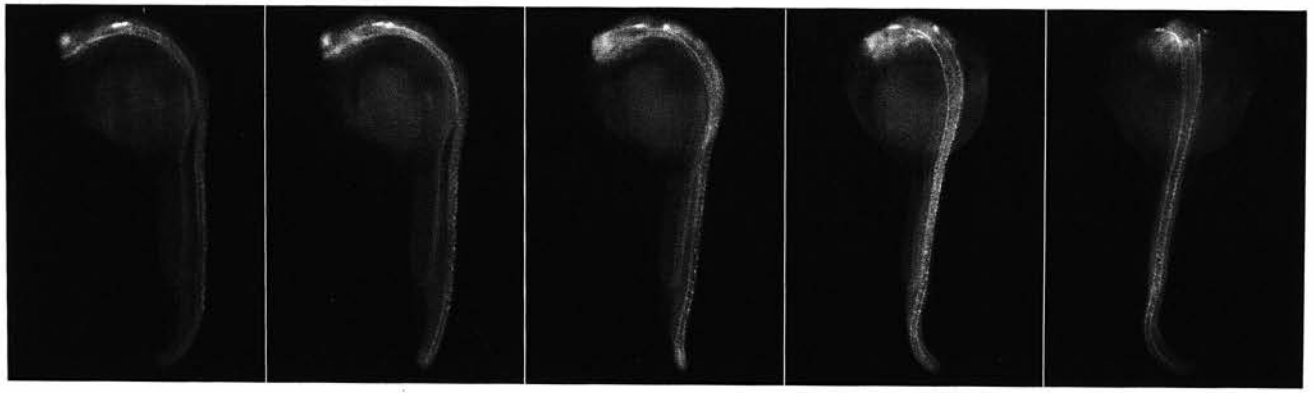


**Fig. 4.4 Images from the scan of the embryo**

Individual images from the rotation of the embryo showing images captured every 22.5° from 0°, a), to 90°, e). The embryo has been labelled fluorescently with an antibody against acetylated tubulin, and background subtraction applied. These two-dimensional images are the projections of the three-dimensional embryo.

The images were converted into wlz format, developed by the Mouse Atlas Project for coding of 3D data, and pixel correction applied to all the images using OptPixelCor (JS).

The high resolution used for the capture of the images revealed a mechanical error in the rotating mechanism of the OPT scanner, that was not evident at the lower magnifications used for mice embryos. During the rotation a repeated lateral movement was observed, such that the sample appeared to be shaking. Uncorrected this would result in a blurred reconstruction so the phase and magnitude of the displacement was determined by careful examination of an orthogonal section of the images (Fig. 4.5), so that the error could be corrected.

From the examination of the orthogonal section it appeared that the displacement occurred in a regular pattern every eight pixels (Fig. 4.6). The motor used to rotate the specimen consisted of four magnets and was being used in half steps and therefore suggested that the error was due to movement of the motor spindle laterally between steps. The magnitude of the displacement was estimated, from the orthogonal section (Fig. 4.6) to be one pixel. In order to correct the error, the images were shifted by one pixel at the maximum displacement and 0.7 at the intermediate steps using WlzAffineTransformObj (MA), the 0.7 pixel shift calculated using linear interpolation. A Perl script, shiftcon (Appendix 2.1) was used to repeat the transformation so it could be applied to all of the images.

Once aligned, the images were processed by the OptAlignProj program (JS). This program calculates the axis of rotation and rotates the images such that the axis of rotation is perpendicular to the y-axis. The output of OptAlignProj is processed by the OptRecon program (JS), which calculates the 3D shape of the object, section by section.

The OptRecon program assumes that the axis of rotation is in the centre of the image, which is not always the case. In order to determine the displacement of the axis of rotation from the centre, a series of individual sections at each end of the

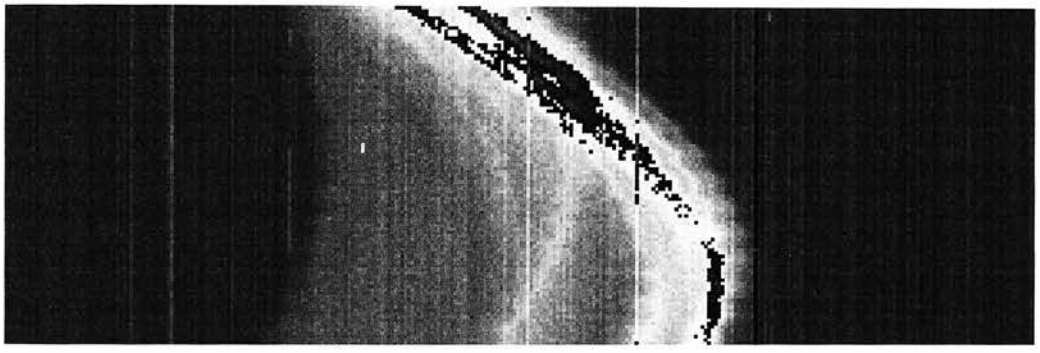


**Fig. 4.5 Orthogonal view of a stack of projection images**

An orthoganal view of a stack of 100 projection images illustrates the repeated lateral movement of the sample. The vertical lines visable on the background are due to differences in the individual sensors of the camera. These lines are removed during background subtraction and by using the OptPixelCor program.

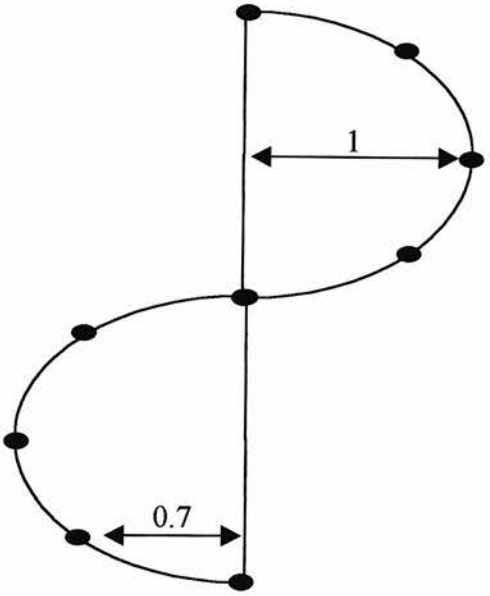



**Fig. 4.6 Correction of Spindle Movement**

The displacement due to the movement of the spindle occurred in a regular pattern with a periodicity of eight. The maximum displacement was estimated to be one pixel. As the displacement followed a sine wave the displacement at the intermediate steps was calculated to be 0.7 (sine 45° X 1, as 1 was the maximum displacement). Four transformations were then applied to the data; +/- 0.7 and +/- 1, with every fourth projection unchanged; to restore the data to a smooth line represented by the straight line in the diagram.

image were reconstructed using OptRecon to test different displacement values. If an incorrect value is used the image will become blurred, and objects start to appear as circles as opposed to discrete spots. The optimum value was determined by examining the reconstruction of individual sections using MA3Dview (MA), testing sections at the top and bottom of the reconstruction. The process of reconstruction is demonstrated in Fig. 4.7 and the supplementary data.

It was observed that a different displacement value was required for the reconstruction of the top and bottom of the images. This shows that despite the correction of the OptAlignProj program, the axis of rotation is still not perpendicular to the images. The angle required to correct the images so that the axis of rotation is perpendicular could be calculated from the difference in displacement value and the distance between measurements. An adjustment of four pixels was required to centre the image at the axis of rotation, as determined by the quality of the section, at section 1000. At section 200 an adjust value of seven was required. Therefore there was a shift of three pixels in 800. Using trigonometry we can calculate the angle of error to be 0.2185 degrees (4dp). Whilst small, this error would account for a displacement of 5.24 pixels between top and bottom of the image.

The images could now be rotated by the angle of error and the full reconstruction carried out using the value of the adjustment at section zero. This would mean that in total three different transformations were applied to the images before the final reconstruction; correction of spindle movement, OptAlignProj correction to make the axis of rotation angle perpendicular to the y-axis, and the calculated rotation angle required to align the axis of rotation. At each transformation information is lost as the data is interpolated during rotation. Therefore it was decided to calculate the single transformation that would produce the same result as the sum of the three transformations and apply this to the pixel corrected images. To correct the spindle movement five different transformations had been applied +/- 1 and +/- 0.7, or a shift of zero. The five different possibilities were then each combined with the two rotations from OptAlignProj and the additional calculated

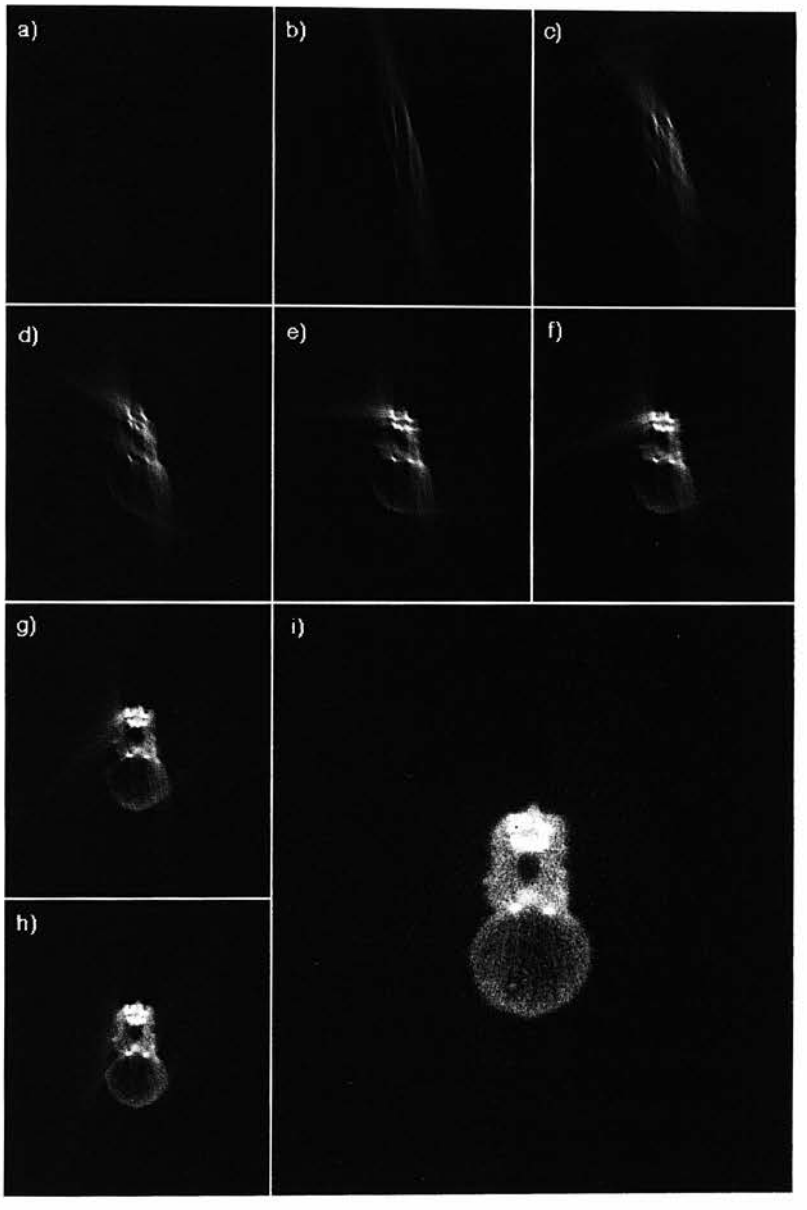



**Fig. 4.7 Reconstruction of a section from the projection images**

The process of reconstruction is shown starting with the reconstruction after 1.8° (a) and then b-h) reconstructions every 22.5° from 22.5° i) reconstruction after 360° when a complete section is produced.

rotation. This resulted in five complete transformations, which describe all of the movements required for each of the projection images. A single transformation (one of the five possible complete transformations) was applied to the images, using the singleshift Perl script (Appendix 2.3), after pixel correction. This removed the need for the OptAlignProj program and the images could be passed straight to OptRecon program, as no other adjustments were necessary.

Due to the large file sizes generated by the high resolution imaging and the large number of angles, the reconstruction was split into a number of smaller reconstructions using a Perl script, remrecon (Appendix 2.4), which split the reconstruction into segments before submitting it to the OptRecon program. This reduced the memory required to compute the reconstruction and also allowed the processing to be split across a number of workstations. Similarly the mrecon script (Appendix 2.5) was used to split the reconstruction into smaller steps for processing on a single workstation.

Once the reconstruction process had been completed, the individual segments were cropped to reduce the image size using MA3Dview and joined to form a single 3D object using WlzUnion (MA). Once assembled into a single object, the grey range was adjusted to remove background and the image was converted to 8-bit using MA3Dview.

The complete process of reconstruction developed in these experiments is demonstrated in Fig. 4.8.

### 4.2.4 Presentation of Reconstructions

The completed reconstruction could now be viewed using MA3Dview. Virtual sections can be taken of the reconstruction at any angle and in any plane. Examples of sections are shown in Fig. 4.9 and the supplementary data.

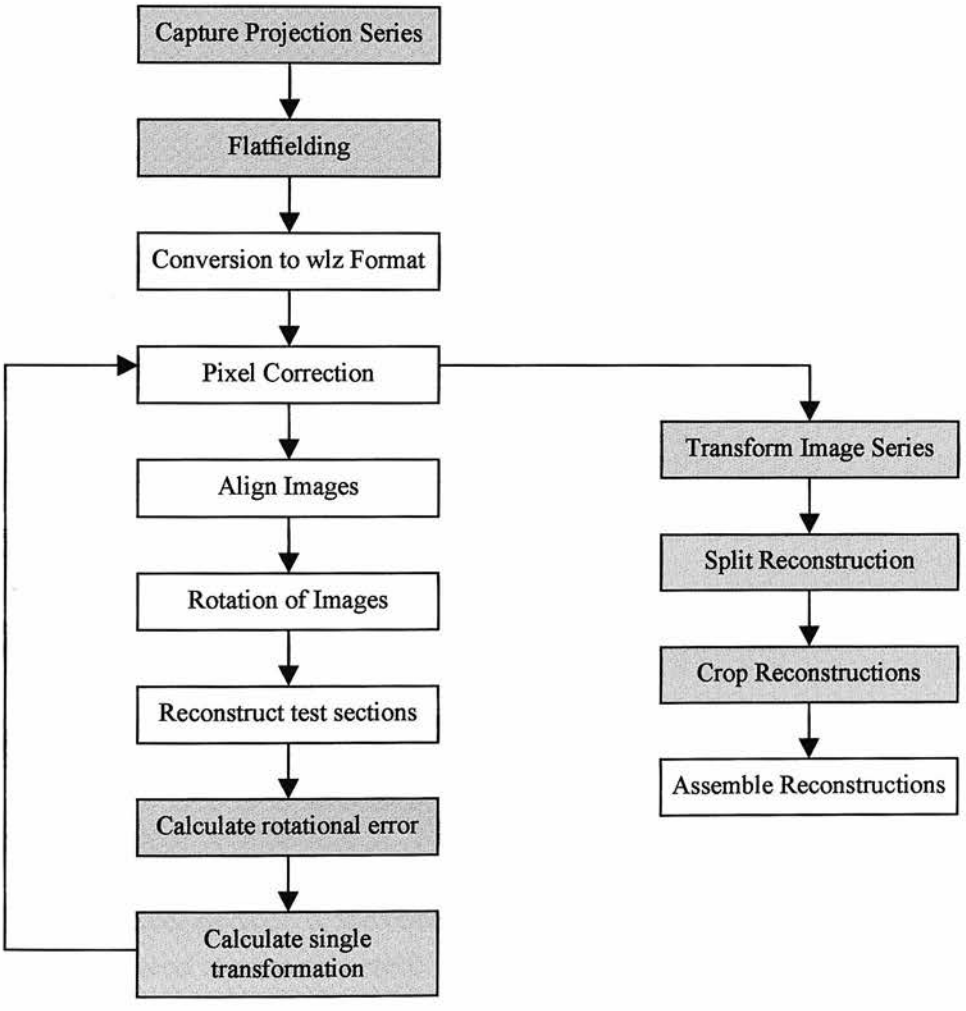



**Fig. 4.8 The zebrafish reconstruction pipeline**

The reconstruction pipeline developed for use on zebrafish embryos. Shaded boxes represent steps that have been created in addition to, or developed from, the standard mouse protocol.

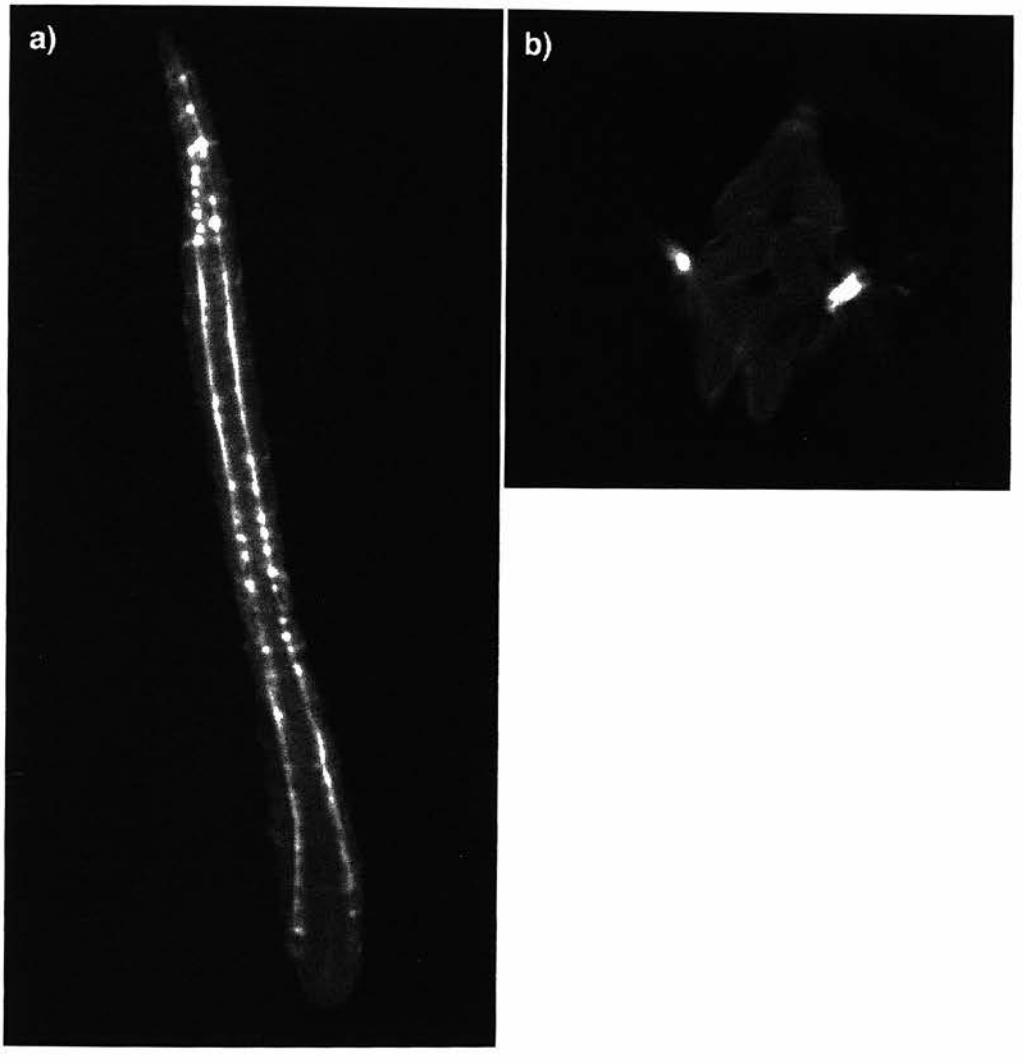



**Fig. 4.9 Virtual sections of the three-dimensional reconstructed object**

The reconstructed object can be virtually sectioned in any plane and at any arbituary angle.
a) a coronal section showing the axon bundles on the dorsal side of the embryo. Individual cell bodies can be distinguished.
b) a horizontal section through the embryo showing very stong expression in the lateral hindbrain

The output from the optical projection tomography is a digital 3D object. One of the advantages of this type of data is that it can be visualised in many ways. As well as the sections shown, we can also use rendering techniques to create iso-surface models or volume models.

Iso-surfaces join values above a set threshold. The visualisation toolkit (VTK, Schroeder et al., 1997) provides tools for the generation and manipulation of surface models and can be used with the OPT reconstructions following conversion to the VTK format using WlzExtFFConvert (MA). By choosing a threshold that includes the fluorescent staining of the embryo we can create a surface model of the entire embryo, this surface model can be viewed in any orientation. By choosing a higher cut off value we can create a surface model of the staining pattern which fluoresces above the level of the PI whole embryo stain, as shown in. Combining both images gives a clear 3D representation of the staining pattern within the context of the embryo, as shown in Fig. 4.10 and the supplementary data.

The VTK also provides tools for generating volume renderings. Volume renderings display all of the data from the reconstruction, and do not use a threshold. The data set is rendered with a reduced opacity so that all of the internal features are still visible compared to surface renderings where a hollow shell representing the shape of embryo is produced. Whilst not necessary for volume rendering, different thresholds can be applied so that the entire embryo or just the staining pattern can be viewed or coloured (Fig. 4.11 and the supplementary data).

## 4.3 Discussion

The results obtained clearly show that OPT can be successfully used to examine embryonic zebrafish antibody staining patterns. The procedure is still at an

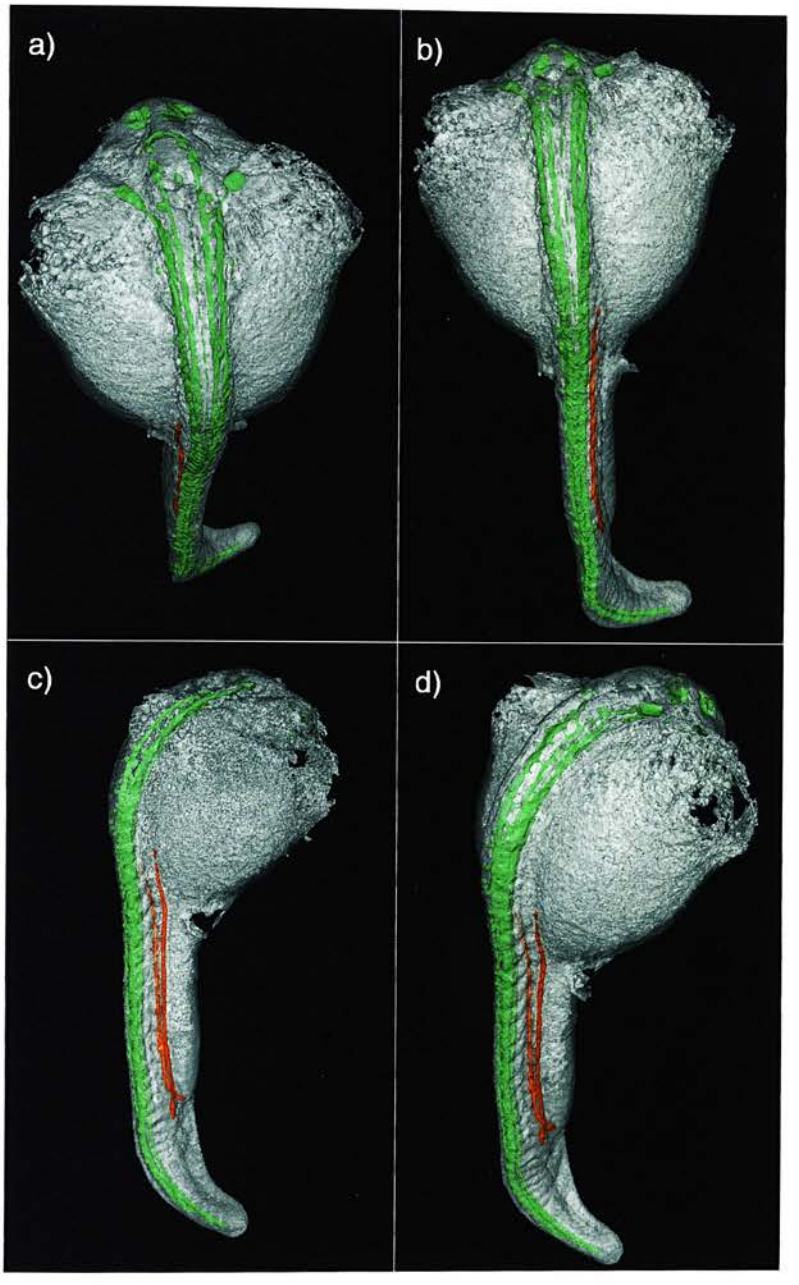



**Fig. 4.10 Surface renderings of reconstruction**

The three dimensional (3D) reconstruction can be volume rendered to give a better visualisation of the 3D object. The rendered object can be rotated to allow any angle to be viewed. Threshold values can be used to identify the antibody signal (green). The high level of blood fluorescence results in the blood being above the threshold and detected as a signal. The blood has been pseudocoloured red.
a-d) A selection of views of the rendered object.

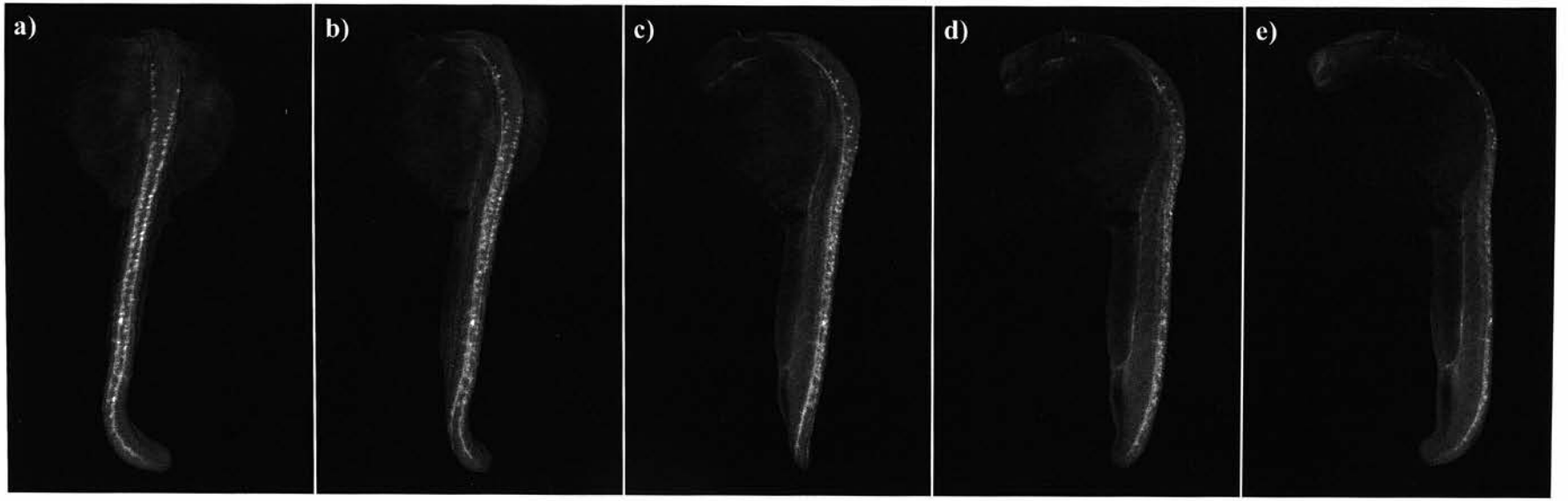



**Fig. 4.11 Volume rendering of the three-dimensional reconstruction**

Volume rendering gives a clear representation of the three-dimensional pattern and a clearer indication of the position of the staining within the embryo than the surface rendering.
a-e) Volume rendering showing a dorsal view and then 22.5° rotations to 90°. The inividual cell bodies and some of the fine projections into the somite can be clearly recognised.

experimental stage and there are a great number of further improvements that could be tested.

### 4.3.1 Sample Preparation

Due to the broad emission spectrum of PI it also emits light in the green range used for the antibody staining. This was used to increase the fluorescence of the embryonic tissue. Experiments with dyes with a narrower emission spectrum, which didn't overlap with the antibody staining, would allow heavier fluorescent staining of the whole embryo and higher quality images. The whole embryo could then be reconstructed separately from the staining pattern and the two reconstructions merged once complete. As well as improving the reconstruction of the embryo, details in the staining pattern would become more evident as background fluorescence is reduced.

We have concentrated on the use of OPT with fluorescent antibody staining but we believe the technique could also be used to characterise expression patterns obtained by in situ hybridisation. The problem with brightfield imaging of in situs stained embryos, the lack of unstained embryonic features, could be solved by taking both a brightfield scan to record the staining and a fluorescent scan to allow the position of the stain within the embryo to be determined. Fluorescent in situs would of course also solve this problem, although the need to clear the embryos in alcohol prevents the use of these techniques at present, as the fluorescent stains used are soluble in alcohol.

### 4.3.2 Resolution

The resolution available is sufficient to capture much of the detail of the fine staining pattern examined, although some of the finer processes can not be individually

resolved and appear as a group of large axons. Whilst resolution would be increased at a higher magnification there would also be further problems created. The magnification used was chosen so as to allow the entire embryo to be imaged. It would be possible to just reconstruct a portion of the embryo at high magnification, although the maximum magnification that could be obtained with the current apparatus would be 63X. Using fluorescent imaging a stain will be visualised as the width of an entire pixel if the emission of light is sufficient to be recorded by the sensor. This is the case even if the cell or cells stained occupy a fraction of the portion of the image that is projected onto the single pixel. At the magnification used to collect the projections, 40X, a single pixel represents 1.2μm, the resolution therefore being 2.4μm. The maximum resolution we could achieve with the current apparatus would be 1.5μm, at 63X magnification. This would be the minimum distance between two lines which would allow them to be distinguished, although as described before any staining less than one pixel width would appear as though it were an entire pixel wide.

As the magnification increases it is more critical that the apparatus be free of vibrations and movements, as these too will be amplified. At the 40X magnification used, movement of the sample was observed and had to be corrected. It would of course produce better results if there were no vibrations and modification to the OPT machinery are being developed as a result of the aberrations discovered in this study. We therefore believe that whilst the magnification could be increased with the current apparatus, the redesign of the apparatus is necessary for significant improvement. The level of magnification that could ultimately be achieved is determined by the depth of focus, which decreases as magnification increases. In order to be successfully reconstructed the sample must be in focus for at least $180^{o}$ and therefore the depth of focus must be at least half the width of the sample.

Increasing the number of angles during the rotation would improve the quality of the results, particularly at the edge of the reconstruction, as the distance between data points would be reduced and the interpolation error decreased accordingly.

### 4.3.3 Reconstruction

Whilst these modifications would improve the resolution possible, it is likely that improvements to the reconstruction software would produce better results with the same data set. The reconstructions described were carried out using a reconstruction grid of the same dimensions as the original images. Images were captured with a width of 1040 pixels. Therefore the reconstructed sections were interpolated onto a grid 1040 pixels square. It is possible to use a larger grid than the original image dimensions for the reconstruction and this may again lead to an improvement in quality, particularly at the centre of rotation where there is less distance between data points (see Fig. 4.3).

The reconstruction process used assumes there is no diffraction of light by the sample. When an embryo is cleared diffraction is greatly reduced, as the different mediums in the embryo are replaced with benzyl benzoate benzyl alcohol. Although diffraction is reduced, it is not eliminated. If the diffraction were uniform it would be possible to reconstruct the embryo using an estimate of the refractive index. However in an embryo there are likely to be many inhomogeneneities and changing refractive indices between tissues. There are approximations that allow reconstructions when there are weak inhomogeneities and it is possible that some of these methods would improve the results of the reconstructions (reviewed in Kak and Slaney, 1988). These methods may also be necessary for higher magnifications, as the diffraction would be more evident.

### 4.3.4 Visualisation

The visualisation of 3D data requires rendering. Both methods of 3D rendering described provide useful information. Surface rendering reduces the size of the data set and allows faster manipulation and display of the images. Generating surfaces however requires a threshold value at which to draw a surface, internal details are

lost and, if boundaries are not well defined, the surface generated may not accurately portray the sample. Volume renderings produce a much larger data set, but all of the information is displayed including internal structures and feint staining that may have been below the threshold for a surface rendering. As the tools for visualisation improve, the interpretation of the data will become easier. Although as the resolution improves and the data sets increase in size, the visualisation tools will have to be developed in order to prevent them from being a limiting factor.

### 4.3.5 Conclusions

At present 3D reconstructions can be generated at very high resolution using confocal microscopy, however, there is a limit to the depth of tissue that can be imaged, and the technique can not be applied to non-fluorescent samples. OPT can be used to study conventional non-fluorescent staining patterns that can not be analysed by confocal studies. Therefore OPT fills a void in current techniques, allowing the generation of 3D data, for samples outside the range of confocal analysis, and the analysis of non-fluorescent data.

The generation of high-resolution 3D models at various stages of development will provide valuable staging series and atlases. As only a single embryo would be required at each stage this would be even more beneficial in the study of organisms where embryonic samples are hard to obtain. With this is mind we have carried out preliminary OPT analysis on dogfish embryos, and hope to develop 3D data sets for a range of fish species.

The large quantities of information that can be extracted from a single embryo means that OPT is suitable for use in high-throughput screens, characterising phenotypes or expression patterns. Recording 3D data from a single embryo will prevent the need for repeated sectioning and improve the speed at which screens can be carried out, and greatly aid the analysis of the results. The digital data can be easily disseminated, allowing more widespread analysis of the results. In addition to its use for

description, OPT also produces quantitative data in the intensity levels of expression. It may be possible to use this data to compare levels of expression between different tissues in a quantitative manner.

We are currently developing methodologies to clear adult zebrafish of pigment to allow the use OPT to study adult anatomy. Due to the size constraints of confocal samples, this would be the highest resolution method of imaging a complete adult fish in 3D. Ultimately we would like to be able to image the development of live embryos, taking advantage of the optical clarity of zebrafish. Currently the diffraction and refraction of light by living tissue makes this impossible, but it is possible that solutions could be developed that would enable us to image live fish. Offering a unique possibility of following development in 3D *in vivo*. Coupled with other techniques such as fluorescent protein expressing transgenics this would be a powerful new tool to study development.

The results of these experiments demonstrate that OPT can be used on samples as small as zebrafish embryos and provide results that are of sufficient quality to provide valuable information about a staining pattern or anatomical features. These developments therefore, allow the application of OPT to a much wider range of organisms and samples than previously possible. The improvements, described here for use on zebrafish embryos, are applicable to all OPT use, and will improve the quality of reconstructions obtained.

We have concentrated on the development of OPT for the analysis of zebrafish and other developmental model organisms. OPT can of course be used on any biological samples that can be treated to allow the passage of light. This means that almost any biological tissue can be examined in this way, offering possibilities in a wide range of areas such as diagnostic medicine and pathology.

We are confident that even higher resolutions will be possible as the technique is developed, and the problems, many of which have apparent solutions (outlined

above), are overcome. Furthermore as OPT becomes more widely available we believe 3D data could become the standard means of presenting results.

## 4.4 Glossary

| | |
|---|---|
| 12-bit | In a 12-bit image every pixel has a value between 0 and 4096 |
| 8-bit | In an 8-bit image every pixel has a value between 0 and 256 |
| Interpolation | Calculation of intermediate values from surrounding known values |
| IPLab | Image capture and manipulation software (Scanalytics) |
| MA3Dview | A program to allow the visualisation of reconstructions in 2D |
| mrecon | A script that splits a reconstruction into smaller segments |
| OptAlignProj | A program that rotates the projection images to make the axis of rotation perpendicular |
| OptPixelCor | A background correction program |
| OptRecon | A program that calculates a 3D reconstruction from projection images |
| Perl | Practical extraction and report language, a programming language |
| Projection | Removal of a dimension from an image |
| Radon transform | A projective transform that produces polar coordinates from a 2D object |
| Ramp filter | A filter which has an effect proportional to the value of the input |
| remrecon | A script that allows mrecon to be run on multiple remote workstations |
| shiftcon | A script that controls the adjustment of projection images to correct for vibrations |
| singleshift | A script that applies single transformation to all of the projection images to prepare them for reconstruction |
| Tomography | The retrieval of information about an object from its sections or projections |
| VTK | The Visualization Toolkit |
| wlz | The image format used for reconstructions |
| WlzAffineTransformObj | A program for the rotation and shifting of image files |
| WlzExtFFConvert | A program that converts a wlz object to other formats |

# Chapter 5: Methods

## 5.1 Solutions

Ampicillin: 1000X 50mg/ml diluted in ethanol

Betagalactosidase staining buffer -1ml: 77µl 0.2M $Na_2HPO_4$, 23µl 0.2M $NaH_2PO_4$, 30µl 5M NaCl, 30µl 0.1M $K_4Fe_3(CN_6)_6$, 30µl 0.1M $K_3Fe_2(CN_6)_6$, $dH_2O$ to 1ml

Denature: 0.5M NaOH, 1.5M NaCl

Half Term: 200mM Tris 9.0, 5mM MgCl, in $dH_2O$.

Kanamycin: 500X 25mg/ml diluted in $dH_2O$

Loading buffer: (30ml $dH_2O$, 5g EDTA, 4.5g Ficoll) pH 8.0, orange G to colour

Luria Agar: 1% Typtone, 0.5% yeast extract, 1% NaCl, 1.5% Agar

Luria Broth: 1% Typtone, 0.5% yeast extract, 1% NaCl, 0.1% Glucose

Neutraliser: 1.5M NaCl, 0.5M Tris (pH7.5)

TAE: 0.8M Tris-HCl pH 8.0, 20mM EDTA, 0.4M acetic acid

λ buffer: 10mM Tris HCl pH 7.5, 10mM $MgSO_4$

## 5.2 Bacterial Culture

### 5.2.1 Transformation

Heat shock transformation was carried out by adding 2µl of DNA to 30µl thawed bacteria (strain DH10B) on ice. The DNA was left for 30 minutes and then the bacteria were heatshocked by immersion in a 42 °C waterbath for 1 minute followed by incubation in ice for 2 minutes. The bacteria were then resuspended in 1ml of Luria broth and incubated at 37 °C for 1 hour in an orbital shaker. The bacteria were then spread on a Luria agar plate containing the required antibiotic for selection under aseptic conditions using an inoculation loop. The agar plates were then inverted and incubated at 37 °C overnight and then stored at 4 °C.

Alternatively electroporation was use to transform bacteria. 2µl of DNA was spotted onto a 25mm 0.025µm filter (Millipore) floating in a Petri dish of $dH_2O$. The DNA was left for at least 30 minutes for dialysis to occur. The DNA was then added to 30µl thawed bacteria on ice (strain XL1-blue) and left for 2 minutes on ice. The Bacteria DNA mix was then moved to an 2mm electroporation cuvette (Equibio) and electroporated using a Biorad Gene pulser set at; 25µFD capacitance, 12V, and 200 ohms. 1ml of Luria Broth was then added to the cuvette and the mixture transferred to a 5ml bijoux tube, it was then moved to 37 °C for 1 hour in an orbital shaker. After 1 hour 100µl, 10µl, and 1µl of the mix was plated onto Luria agar plates containing the antibiotic required for selection. For the 10µl and 1µl plates the mixture was made up to a volume of 100µl using Luria Broth. The plates were then incubated overnight at 37 °C.

### 5.2.2 Bacterial Cell Culture

Bacteria were streaked onto Luria Agar plates containing the required antibiotic. The plates were then incubated overnight at 37°C. The following morning the plates were wrapped in parafilm to prevent condensation and moved to 4°C, the plates were stored at 4 °C for up to 2 weeks.

For DNA isolation a single colony was collected from the plate using a pipette tip and transferred to a Luria Broth (LB) solution in a vessel at least 3 times the volume of the solution. The volume of LB was 10ml for miniprep, 50ml for midipreps, and 100ml for maxi preps. The solution was then left shaking at 37°C, 250 rpm, for 16 hours. The following morning the DNA was isolated using Qiagen spin miniprep kit, Qiagen hi-speed midiprep kit, or Qiagen maxiprep kit. All kits were used according to the manufacturer's instructions excluding the optional additional steps.

## 5.3 DNA Manipulation

### 5.3.1 Restriction Digests

Digests were carried out using Roche or New England Biolabs enzymes and buffers. Reactions were carried out in 20µl using 1µl (10 units) of the required enzyme. When 2 different enzymes were used the optimal buffer was chose using the manufacturers recommendation and the reaction volume increased to 30µl to reduce the final glycerol concentration. The restriction digests were incubated in a 37°C waterbath for 2 hours or overnight.

### 5.3.2 Phenol Chloroform Extraction

The volume of the DNA or RNA solution was made up to 100µl with $dH_2O$ and an equal volume of buffer saturated phenol added. The solutions were mixed and centrifuged at 13000rpm for 15 seconds. The top, aqueous, layer was then removed. A further 100µl of $dH_2O$ was added to the phenol and the procedure repeated. 200µl of chloroform isoamyl alcohol (24:1) was added to the solution, extracted from the phenol, mixed and centrifuged, again the top, aqueous layer was removed and added to a fresh eppendorf. 200µl of $dH_2O$ was added to the chloroform and the procedure repeated. The final solution was then ethanol precipitated.

### 5.3.3 Ethanol Precipitation

To the 400µl DNA or RNA solution 1ml of ethanol and 40µl of 3M NaAc pH5.2 (DNA) or 40µl 4M LiCl (RNA) was added and the mixture was then stored at -20 °C at least 1 hour. The DNA or RNA was pelleted by centrifugation at 13000 rpm for 10 minutes (4 °C for RNA) and resuspended in $dH_2O$ (DNA) or DEPC treated $dH_2O$ (RNA).

### 5.3.4 Agarose Gel Electrophoresis

Unless otherwise stated all gels were 1% agarose gels containing 2µl of Ethidium Bromide (10mg/ml Sigma) made and run with 1 X TAE. 10% loading buffer was added to each sample before loading. Molecular weight markers were run alongside DNA samples, and were either molecular weight marker X (Roche), log-2 (NEB), molecular weight marker III (Roche), or ϕ174 (Promega). The markers were chosen depending on the expected size of the product.

### 5.3.5 Gel Extraction of DNA

DNA bands were excised using a scalpel and then purified using either the Geneclean II kit (BIO101) or the Qiagen gel extraction spin kit. Both kits were used according to manufacturer's instructions with the exception that $dH_2O$ was used as the eluate, and no difference in quality was observed between the two kits. For suspension in small volumes, less than 30 µl, the Geneclean II kit was used otherwise the Qiagen kit was used.

### 5.3.6 DNA Ligation

Ligations were carried out in a total volume of 5µl containing 0.5µl ligation buffer and 0.5µl T4 DNA ligase (1U/µl Roche). Insert and dephosphorylated vector DNA was added in the ration 3:1 for cohesive end ligation and 6:1 for blunt end ligation. The reaction was left overnight at 16 °C.

### 5.3.7 5' Dephosphorylation

DNA to be dephosphorylated was made up to a volume of 89µl to which 10ul alkaline phosphatase 10X buffer and 1µl of alkaline phosphatase (Roche) was added. The mixture was then incubated for 30 minutes at 37 °C. The dephosphorylated DNA was then recovered by phenol chloroform extraction.

## 5.4 Nucleic Acid Synthesis

### 5.4.1 Capped RNA Synthesis

A reaction mixture was set up containing 5µg linearised DNA, 5µl 10X transcription buffer (Roche), 5µl DTT, 2µl 25X dNTP mix (250mM ATP, 250mM CTP, 250mM UTP, 25mM GTP), 5µl 5mM GpppG (Roche), 2µl RNAse inhibitor (40U/µl, Roche), 2.5µl polymerase (either SP6, T7, or T3, 20U/µl Roche). The reaction mixture was mixed and spun before incubation in a 37°C waterbath for 30 minutes. 2.5µl 10mM GTP was added and the mixture incubated for an hour at 37°C. After 1 hour 5µl DNAseI was added (10U/µl Roche) and the mixture incubated for 30 minutes at 37 °C for 30 minutes.

The RNA was then phenol chloroform extracted and ethanol precipitated.

### 5.4.2 Sequencing

A reaction mixture was set up containing 4µl of rhodomine sequencing mix (Applied Biosystems), 4µl half term, 1µl of primer (0.1µg/µl). To this mix denatured DNA was added depending on the sequencing template, 500ng for plasmids and 1µg for cosmids. The DNA was denatured by incubation at 96 °C for 5 minutes followed by incubation on ice. The volume of the mixture was then made up to 20µl using $dH_2O$. The sequencing was then carried out using a GRI dyad with the following program. 96 °C 30s, 50 °C 30s, 60 °C 4 minutes, this program was then repeated 29 times.

Once the reaction was complete the DNA was precipitated. 50µl of ETOH and 2µl of 3M NaAc, and 0.5µl pellet paint Co-Precipitant (Novagen) was added to the reaction

and incubated on ice for 30 minutes. The mixture was then centrifuged at 13,000rpm for 30 minutes and the supernatant removed. The pellet was then allowed to air dry before the tube was sealed.

The sample was then submitted to the sequencing service where it was loaded and run on an ABI Prism sequencer.

### 5.4.3 Oligonucleotide Design and Synthesis

Oligonucleotides were designed using the primer3 software at http://www-genome.wi.mit.edu/cgi-bin/primer/primer3_www.cgi and synthesised by either MWG biotech or Genosys. A table of oligonucleotides used is presented in the Appendix 1.1.

### 5.4.4 Polymerase Chain Reaction

A reaction mix was made containing 5µl PCR buffer, 2µl $MgCl_2$ (25mM, Roche), 1µl dNTPs (25X dNTP mix made containing 10mM dATP, dCTP, dGTP, and dTTP, ABGene), 1µl of each primer (0.1µg/µl), 0.1µl Taq polymerase (5U/µl Roche), and $dH_2O$ to take the volume to 50µl minus the volume of template needed (approximately 25ng), per reaction. The reaction was then aliquoted and the template added, or $dH_2O$ for the negative control. The mixture was then mixed and briefly centrifuged before being placed in the PCR machine. Reaction were carried out in 200µl tubes when using the GRI dyad PCR machines or 500µl tubes and 2 drops of mineral oil added to the top of each reaction when using the Hybaid omnigene PCR machine.

The PCR program used was 94 °C for 5 minutes and then 30 cycles of 94 °C for 30 seconds, 55 °C for 30 seconds, 72 °C for 1 minute, followed by 10 minutes at 72 °C.

The elongation time was increased by 1 minute for every 1kb of the expected product. The annealing temperature was reduced to 5 °C below the annealing temperature of the primers if less than 55 °C. Reactions were first tried with the 2mM $MgCl_2$ shown but if optimisation was required 1mM, 1.5mM, 2mM, 2.5mM, and 3mM, were tried, The negative control was always carried out using 2mM.

### 5.4.5 Morpholinos

Morpholino oligonucleotides were designed and synthesised by Gene-Tools from the 75bp 5' and 25bp 3' of the ATG transcriptional start site. Morpholinos were diluted in $dH_2O$ and injected at a starting concentration of 0.5mM.

## 5.5 Imaging and Immunohistochemistry

### 5.5.1 Wholemount Antibody Staining

Embryos were fixed for 4 hours at room temperature or overnight at 4 °C. The embryos were then washed in 0.1M $PO_4$ buffer for 5 minutes, followed by $dH_2O$ for 5 minutes. The water was then removed and replaced with acetone and incubated at -20 °C for 7 minutes. The acetone was then removed and slowly replaced with water and then left to wash for 5 minutes, followed by a 5 minute wash in 0.1M $PO_4$ buffer. The embryos were then blocked in a mixture containing 2% serum of the secondary antibody host, 1% BSA, 1% DMSO, in PBS for at least 30 minutes. The blocking solution was then replaced with fresh blocking solution containing the primary antibody and left overnight at 4 °C.

The embryos were then washed 8 X 15 minutes with PBS containing 1%BSA and 1% DMSO. The PBS/BSA/DMSO solution was then replaced with a PBS/BSA/DMSO solution containing the secondary antibody and the embryos incubated overnight at 4 °C.

The embryos were then washed 6 X 10 minutes in 0.1M $PO_4$ buffer containing 1% DMSO.

If the secondary antibodies were coupled to fluorophores the embryos were then mounted and imaged as described later (5.5.5 and 5.6). For brightfield analysis a secondary antibody coupled to peroxidase was used. The stain was then developed using DAB peroxidase substrate kit (Vector laboratories inc.) according to the manufacturers instructions

### 5.5.2 In Situ Hybridisation

In situ probe synthesis and hybridisation was carried out as described by Jowett (Jowett, 1998) with the exception that the PBT sheep serum blocking solution was replaced with MAB containing 2% blocking reagent (Roche) and embryos were washed with MAB prior to blocking.

### 5.5.3 Betagalactosidase Staining of *LacZ* Injected Embryos

Embryos were fixed in 4% PFA for 30 minutes at room temperature. The embryos were then washed 5 X 5 minutes with PBST at room temperature and then moved to the dark were they were rinsed for 5 minutes in staining buffer, which was then replaced with staining buffer containing 13.25µl/ml of 50mg/ml Xgal in DMSO. This reaction was then allowed to proceed in the dark at room temperature until blue staining was visible, generally between 15minutes and 4 hours. Once stained the

embryos were washed with PBST 5 X 5 minutes before fixation for 4 hours at room temperature or overnight at 4°C.

### 5.5.4 Propidium Iodide Staining

2µl of propidium iodide (0.1mg/ml) was added to embryos in 1ml of PBST. The embryos were shaken for 1 hour at room temperature and then washed 8 X 5minutes in PBST.

### 5.5.5 Mounting Embryos for Microscopy

Stained embryos were washed in 50% EtOH, 100% EtOH and then 1:2Benzyl Benzoate:Benzyl Alcohol for at least 5 minutes at each step. The embryos were then mounted in DPEX. A bridge was made of coverslips to prevent damage to the embryo using 2 No. 1 coverslips for embryos up to 24hr and 3 after 24hr. For tail mounts no bridges were used.

For confocal microscopy a glycerol series was used of 30%, 50%, and 70% with at least 10 minutes at each stage and the tail then mounted in 70% glycerol without a bridge and covered with a No. 0 coverslip.

For tail mounts the head and yolk were removed using a pair of 30 gauge needles at the 50% EtOH or 70% glycerol stage.

## 5.6 Microscopy

### 5.6.1 Stereo Microscopy

A Leica Fluo III stereomicroscope was used for fluorescent and brightfield imaging of embryos. Images were captured using an RS Photometrics Coolsnap digital camera and IPLab software (version 3.2 and 3.6).

### 5.6.2 Compound Microscopy

A Zeiss Axioskop was used for fluorescent and DIC image capture using either a Hamamatsu CA742-95 digital camera or a Zeiss controlled film camera with Fuji 64T slide film. The capture of digital images and the control of filter wheels was carried out using a Macintosh computer running IPLab (versions 3.2 and 3.6).

### 5.6.3 Confocal Microscopy

Confocal microscopy was carried out on either a Zeiss LSM 510 confocal microscope and images captured through the LSM software on a PC or a Biorad MRC 600 operated using CoMOS (version 7.0a) software on a PC. Image quality was greatly improved using the LSM 510 and this was used when possible.

## 5.7 Library Screening and Hybridisation

### 5.7.1 Plating λ Libraries

Approximately 200,000 pfu of phage were plated onto each of four 22cm square plates. The phage was mixed with 1ml of plating bacteria and incubated for 15 minutes at 37°C, to allow adsorption of phage to host, then this was mixed with 30ml CY-top containing 10mM $MgSO_4$ at 50 °C. This was plated onto warmed and dried L-Agar plates. The library plates were then incubated overnight at 37 °C.

Phage DNA was lifted onto nylon filters (Hybond-N Amersham Pharmacia Biotech). The filters were laid onto the plates for 2 minutes and the orientation marked using a needle dipped in Indian ink. The filters were then placed in denature solution for 4 minutes, neutraliser for 4 minutes and 2XSSC for 30 seconds. Filters were crosslinked by UV irradiation using a XL-1500 UV crosslinker (Spectronics Corporation) at the optimum crosslink setting, and then baked in an 80 °C oven for 2 hours. The dried membrane was then stored between two sheets of Whatman paper until use.

Hybridisation of the filters was then carried out as described in (5.7.4).

Positive plaques were excised from the plate using the large end of a sterile 1ml Gilson tip. The plug was then placed into 1ml of λ buffer and eluted for at least 30 minutes at room temperature.

For secondary screening a 1:100 dilution was made of the eluted plug in λ buffer and 5µl, 10µl, and 20µl of the dilution was mixed with 100µl XL-blue bacteria in 10mM $MgSO_4$ and incubated at 37 °C for 15 minutes. The bacteria and phage were then plated in 3ml of molten (50 °C) CY-top onto 90mm L-agar $Mg^{2+}$ plates and incubated

overnight at 37 °C. For the secondary screening the phage was lifted onto 82mm Protran discs (Schleicher and Schuell). The filters were then hybridised as previously described.

Once positive clones had been identified they were excised using ExAssist helper phage with SOLR strain (Stratagene) according to the manufacturers instructions.

### 5.7.2 Southern Blotting

Once photographed the gel was gel was immersed in 2.5 times its volume of 0.25M HCl and shaken for 20 minutes to fragment the DNA. The gel was then rinsed in $dH_2O$ for 20 minutes. The rinsed gel was then washed twice in denaturant for 20 minutes followed by 40 minutes in neutraliser.

The blotting apparatus was set up as follows. 2 sheets of Whatman 17Chr paper were laid on a plastic sheet across a tray containing 20X SSC such that the ends of the Whatman paper was immersed in the 20XSSC. The gel was then laid onto the Whatman, the air bubbles carefully removed, and the bottom right corner cut to mark the orientation. The paper and tray was then covered with saran wrap up to the edge of the gel to prevent evaporation The hybridisation membrane, Hybond-N (Amersham Pharmacia Biotech) was then laid on top of the gel and a glass pipette used to remove the air bubbles. A sheet of Whatman 17Chr paper larger than the gel was then laid on top of the membrane and approximately 5cm of paper towels placed onto the paper. A sheet of glass was then laid across the paper towels and a 250ml bottle placed on top. The gel was the allowed to blot for 24 hours before being washed for 5 minutes in 20X SSC. The washed gel was then crosslinked by UV irradiation using a XL-1500 UV crosslinker (Spectronics Corporation) at the optimum crosslink setting, and then baked in an 80 °C oven for 2 hours. The dried membrane was then stored between two sheets of Whatman paper until use.

### 5.7.3 Radioactive Labelling of Probes

50ng of DNA was made up to 12µl with $dH_2O$ and heated at 96 °C for 10 minutes and then moved immediately to ice. After 5 minutes on ice 5µl of Hi-Prime (Roche) followed by 3µl of $^{32}P\alpha CTP$ (Amersham Pharmacia Biotech). The mixture was then incubated at 37°C for 1 hour.

A Sephadex G-50 DNA grade NICK column (Amersham Pharmacia Biotech) was equilibrated using 400µl of TE. The reaction mixture was made up to a volume of 100µl using TE. Once the column had equilibrated the reaction mixture was added followed by an additional 300µl of TE and the filtrate discarded. A further 400µl of TE was added and the elute collected.

### 5.7.4 Hybridisation of Radiolabelled Probes

Filters were placed in Hybaid glass bottles, up to 3 filters were placed in a single bottle separated by gauss. 20ml of prehybridisation solution was added to each bottle and the filters were then incubated at 65 °C in a rotating oven for at least 2 hours. The prehybridisation solution was then replaced with 20ml of hybridisation solution and 200υl sonicated salmon sperm (50mg/ml, Sigma) was added together with the radiolabelled probe. The filters were then incubated overnight at 65 °C in the rotating oven. The hybridisation mixture was then removed and the filters washed with 2XSSC 0.1% Tween 3 times at in the bottles at 65 °C for 30 minutes. The filters were then removed from the bottles and washed in a tray at 65 °C. For cross species hybridisation the filters were washed 3 times with 2XSSC 0.1%Tween for 30 minutes, for same species hybridisation the stringency of the wash was increased to 0.2XSSC 0.1% Tween. The filters were then individually sealed in polythene bags. The filters were than placed in a film cassette and the orientation of the film marked by fluorescent labels. Autoradiography film, either Biomax or x-omat (Kodak), was added to the cassettes and the cassettes were then stored at -80 °C. The length of

exposure was determined by the strength of the counts on the filters and varied from 20 minutes to 1 week.

### 5.7.5 Libraries

Grided libraries; ZF early cDNA #717 (RZPD), ZF liver cDNA #532 (RZPD), ZF 26S cDNA #39 (RZPD).

Phage library Lambda Zap II 15-19hr polyA+ (Stratagene).

## 5.8 Computational Methods

### 5.8.1 Programs

| Program | version | Reference |
|---|---|---|
| BLAST | 2.1.3 | (Altschul et al., 1997) |
| Clustalw | 1.74 | (Thompson et al., 1994) |
| Consed | 11.0 | (Gordon et al., 1998) |
| Dotter | 3.0 | (Sonnhammer and Durbin, 1995) |
| ExonPlot | 1.0 | Taylor, unpublished |
| ForCon | | (Raes and Van de Peer, 1999) |
| Mega2 | 2.1 | (Kumar et al., 2001) |
| needle | 22/10/2000 | Bleasby, unpublished |
| Phrap | | Ewing and Green, unpublished |
| Phred | 0.000925.c | (Ewing et al., 1998; Ewing and Green, 1998) |
| NIX | | Williams, unpublished http://www.hgmp.mrc.ac.uk/NIX/ |
| RepeatMasker | 07/07/2001 | Smit, unpublished |
| StrataSplice | 2001 | Levine and Durbin, unpublished |
| wise2 | 2-1-22c | Birney, unpublished |

### 5.8.2 Databases

| Database | Content | Location |
|---|---|---|
| EST | nucleotide | http://www.hgmp.mrc.ac.uk |
| Fugu genome shotgun | nucleotide | ftp://ftp.jgi-psf.org/pub/JGI_data/Fugu |
| Fugu genome assembly | nucleotide | ftp://ftp.jgi-psf.org/pub/JGI_data/Fugu |
| SPTR | protein | http://www.ebi.ac.uk/swissprot/ |
| Tetraodon genome shotgun | nucleotide | http://www.genoscope.cns.fr/externe/tetraodon |
| WashU-Zebrafish Genome Resources | Clustered ESTs and nucleotide | http://zfish.wustl.edu/ |
| Zebrafish genome shotgun | nucleotide | http://trace.ensembl.org and HGMP project-blast |

### 5.8.3 Genomic Sequence Assembly

Initially an indexed BLAST database was created from the publicly available sequence using the formatdb application within the blast package. The indexed database was then searched with the protein or nucleotide sequence and the sequences of the significant hits retrieved. Where there were many hits only the top hits were retrieved, as the remainder were assumed to be repetitive sequences given the low coverage of the sequencing project. Once the sequences has been retrieved they were assembled using PhredPhrap and viewed using consed. The contigs created were compared to the original query sequence and matching contigs were then used to search the database again by BLAST and the process repeated in order to extend the contigs.

This process was later carried out using the zebedee program (Semple and Bryson-Richardson, Appendix 2.6). The program automated this process described above

with the difference that BLAST results were parsed by identity instead of significance to try and improve the identification of homologous sequences.

### 5.8.4 Gene Predictions

Coding sequences was predicted from genomic sequence using the genewise package. An alignment of the known amino acid sequences was aligned using the clustalw program and the output used to create a Hidden Markov Model (HMM) using the hmmbuild program. Predictions were then made using the genewise program, as the program is optimised for human splice sequences, predictions were made using the -splice flat and -intron tied function to improve prediction of splice sites.

### 5.8.5 Phylogenetic Analysis

Alignments were created using clustalw. The output of the alignment was then converted to Mega2 format using the ForCon program. Mega2 was then used to construct and bootstrap the phylogeny. Minimum evolution with complete deletion was used to construct the phylogeny. Bootstrapping was carried out with 1000 repetitions with a random seed.

## 5.9 Microinjection and Zebrafish Strains

### 5.9.1 Microinjection

An injection mixture was made consisting of either 0.1μg/μl DNA or RNA, or 0.5mM morpholino, and 1% phenol red in $dH_2O$. These starting concentrations were then adjusted depending on the results observed.

A needle was broken near the tip using forceps. 2μl of injection was spotted onto a microscope slide and the needle filled using the fill function on a Narishige IM 300 microinjector. The tip of the needle was then adjusted using forceps such that it passed easily through the chorion.

Needle used were 1mm boro-silicate capillaries (1mm outer diameter, 0.78 inner diameter, Clark electromedical instruments) pulled on a CFP micro-electrode puller.

### 5.9.2 Zebrafish Strains

In situ analysis, immunohistochemistry and microinjections were carried out using the golden mutant line due to its reduction of pigment. Some embryos of the Wik, AB, and Tu strains were also used at stages less than 26-somites for in situ and immunohistochemistry

# Bibliography

Amores A., Force A., Yan Y. L., Joly L., Amemiya C., Fritz A., Ho R. K., Langeland J., Prince V., Wang Y. L., Westerfield M., Ekker M., and Postlethwait J. H. (1998) Zebrafish hox clusters and vertebrate genome evolution. *Science* **282,** 1711-1714.

Atomi Y., Yamada S., and Nishida T. (1991) Early changes of alpha B-crystallin mRNA in rat skeletal muscle to mechanical tension and denervation. *Biochem Biophys Res Commun* **181,** 1323-1330.

Atomi Y., Toro K., Masuda T., and Hatta H. (2000) Fiber-type-specific alphaB-crystallin distribution and its shifts with T(3) and PTU treatments in rat hindlimb muscles. *J Appl Physiol* **88,** 1355-1364.

Banerjee U., Renfranz P. J., Pollock J. A., and Benzer S. (1987) Molecular characterization and expression of sevenless, a gene involved in neuronal pattern formation in the Drosophila eye. *Cell* **49,** 281-291.

Banerjee-Basu S., Landsman D., and Baxevanis A. D. (1999) Threading analysis of prospero-type homeodomains. *In Silico Biol* **1,** 163-173.

Barresi M. J., D'Angelo J. A., Hernandez L. P., and Devoto S. H. (2001) Distinct mechanisms regulate slow-muscle development. *Curr Biol* **11,** 1432-1438.

Belecky-Adams T., Tomarev S., Li H. S., Ploder L., McInnes R. R., Sundin O., and Adler R. (1997) Pax-6, Prox 1, and Chx10 homeobox gene expression correlates with phenotypic fate of retinal precursor cells. *Invest Ophthalmol Vis Sci* **38,** 1293-1303.

Benjamin I. J., Shelton J., Garry D. J., and Richardson J. A. (1997) Temporospatial expression of the small HSP/alpha B-crystallin in cardiac and skeletal muscle during mouse development. *Dev Dyn* **208,** 75-84.

Bier E., Vaessin H., Younger-Shepherd S., Jan L. Y., and Jan Y. N. (1992) deadpan, an essential pan-neural gene in Drosophila, encodes a helix-loop-helix protein similar to the hairy gene product. *Genes Dev* **6,** 2137-2151.

Blagden C. S., Currie P. D., Ingham P. W., and Hughes S. M. (1997) Notochord induction of zebrafish slow muscle mediated by Sonic hedgehog. *Genes Dev* **11,** 2163-2175.

Blake D. J., Love D. R., Tinsley J., Morris G. E., Turley H., Gatter K., Dickson G., Edwards Y. H., and Davies K. E. (1992) Characterization of a 4.8kb transcript from the Duchenne muscular dystrophy locus expressed in Schwannoma cells. *Hum Mol Genet* **1,** 103-109.

Bodmer R., Carretto R., and Jan Y. N. (1989) Neurogenesis of the peripheral nervous system in Drosophila embryos: DNA replication patterns and cell lineages. *Neuron* **3,** 21-32.

Borycki A. G., Mendham L., and Emerson C. P. J. (1998) Control of somite patterning by Sonic hedgehog and its downstream signal response genes. *Development* **125,** 777-790.

Brady J. P., Garland D. L., Green D. E., Tamm E. R., Giblin F. J., and Wawrousek E. F. (2001) AlphaB-crystallin in lens development and muscle integrity: a gene knockout approach. *Invest Ophthalmol Vis Sci* **42,** 2924-2934.

Broadie K. and Bate M. (1993) Muscle development is independent of innervation during Drosophila embryogenesis. *Development* **119,** 533-543.

Broadus J., Fuerstenberg S., and Doe C. Q. (1998) Staufen-dependent localization of prospero mRNA contributes to neuroblast daughter-cell fate. *Nature* **391,** 792-795.

Bruce A. E., Oates A. C., Prince V. E., and Ho R. K. (2001) Additional hox clusters in the zebrafish: divergent expression patterns belie equivalent activities of duplicate hoxB5 genes. *Evol Dev* **3,** 127-144.

Brune R. M., Bard J. B., Dubreuil C., Guest E., Hill W., Kaufman M., Stark M., Davidson D., and Baldock R. A. (1999) A three-dimensional model of the mouse at embryonic day 9. *Dev Biol* **216,** 457-468.

Burglin T. R. (1994) A Caenorhabditis elegans prospero homologue defines a novel domain [letter]. *Trends Biochem Sci* **19,** 70-71.

Campos-Ortega, J. A., Jurgens, G, and Hofbauer, A. (1979) Cell clones and petern formatio: studies on *sevenless*, a mutant of *Drosophila melanogaster. Roux's Arch Dev Biol* **186,** 27-50.

Carthew R. W. and Rubin G. M. (1990) seven in absentia, a gene required for specification of R7 cell fate in the Drosophila eye. *Cell* **63,** 561-577.

Chen W., Burgess S., and Hopkins N. (2001) Analysis of the zebrafish smoothened mutant reveals conserved and divergent functions of hedgehog activity. *Development* **128,** 2385-2396.

Chu-LaGraff Q., Wright D. M., McNeil L. K., and Doe C. Q. (1991) The prospero gene encodes a divergent homeodomain protein that controls neuronal identity in Drosophila. *Development Suppl* **2,** 79-85.

Copertino D. W. and Hallick R. B. (1991) Group II twintron: an intron within an intron in a chloroplast cytochrome b-559 gene. *EMBO J* **10,** 433-442.

Currie P. D. and Ingham P. W. (1996) Induction of a specific muscle cell type by a hedgehog-like protein in zebrafish. *Nature* **382,** 452-455.

Currie P. D. and Ingham P. W. (2001) Induction and Patterning of the Embryonic Skeletal Muscle Cells in The Zebrafish, in *Muscle Development and Growth* (Johnston I. A., ed.), pp. 1-17. Academic Press.

Daga A., Karlovich C. A., Dumstrei K., and Banerjee U. (1996) Patterning of cells in the Drosophila eye by Lozenge, which shares homologous domains with AML1. *Genes Dev* **10,** 1194-1205.

Demidenko Z., Badenhorst P., Jones T., Bi X., and Mortin M. A. (2001) Regulated nuclear export of the homeodomain transcription factor Prospero. *Development* **128,** 1359-1367.

Devoto S. H., Melancon E., Eisen J. S., and Westerfield M. (1996) Identification of separate slow and fast muscle precursor cells in vivo, prior to somite formation. *Development* **122,** 3371-3380.

Doe C. Q., Hiromi Y., Gehring W. J., and Goodman C. S. (1988a) Expression and function of the segmentation gene fushi tarazu during Drosophila neurogenesis. *Science* **239,** 170-175.

Doe C. Q., Smouse D., and Goodman C. S. (1988b) Control of neuronal fate by the Drosophila segmentation gene even-skipped. *Nature* **333,** 376-378.

Doe C. Q., Chu-LaGraff Q., Wright D. M., and Scott M. P. (1991) The prospero gene specifies cell fates in the Drosophila central nervous system. *Cell* **65,** 451-464.

Driever W., Solnica-Krezel L., Schier A. F., Neuhauss S. C., Malicki J., Stemple D. L., Stainier D. Y., Zwartkruis F., Abdelilah S., Rangini Z., Belak J., and Boggs C. (1996) A genetic screen for mutations affecting embryogenesis in zebrafish. *Development* **123,** 37-46.

Du S. J., Devoto S. H., Westerfield M., and Moon R. T. (1997) Positive and negative regulation of muscle cell identity by members of the hedgehog and TGF-beta gene families. *J Cell Biol* **139,** 145-156.

Dubin R. A., Gopal-Srivastava R., Wawrousek E. F., and Piatigorsky J. (1991) Expression of the murine alpha B-crystallin gene in lens and skeletal muscle: identification of a muscle-preferred enhancer. *Mol Cell Biol* **11,** 4340-4349.

Duncan M. K., Cui W., Oh D. J., and Tomarev S. I. (2002) Prox1 is differentially localized during lens development. *Mech Dev* **112,** 195-198.

Ekker M., Wegner J., Akimenko M. A., and Westerfield M. (1992) Coordinate embryonic expression of three zebrafish engrailed genes. *Development* **116,** 1001-1010.

Felsenfeld A. L., Curry M., and Kimmel C. B. (1991) The fub-1 mutation blocks initial myofibril formation in zebrafish muscle pioneer cells. *Dev Biol* **148,** 23-30.

Ferrandon D., Elphick L., Nusslein-Volhard C., and St Johnston D. (1994) Staufen protein associates with the 3'UTR of bicoid mRNA to form particles that move in a microtubule-dependent manner. *Cell* **79,** 1221-1232.

Flores G. V., Daga A., Kalhor H. R., and Banerjee U. (1998) Lozenge is expressed in pluripotent precursor cells and patterns multiple cell types in the Drosophila eye through the control of cell-specific transcription factors. *Development* **125,** 3681-3687.

Fortini M. E., Simon M. A., and Rubin G. M. (1992) Signalling by the sevenless protein tyrosine kinase is mimicked by Ras1 activation. *Nature* **355,** 559-561.

Freeman M. (1996) Reiterative use of the EGF receptor triggers differentiation of all cell types in the Drosophila eye. *Cell* **87,** 651-660.

Glasgow E. and Tomarev S. I. (1998) Restricted expression of the homeobox gene prox 1 in developing zebrafish. *Mech Dev* **76,** 175-178.

Gonzalez F., Romani S., Cubas P., Modolell J., and Campuzano S. (1989) Molecular analysis of the asense gene, a member of the achaete-scute complex of Drosophila melanogaster, and its novel role in optic lobe development. *EMBO J* **8,** 3553-3562.

Gopal-Srivastava R. and Piatigorsky J. (1993) The murine alpha B-crystallin/small heat shock protein enhancer: identification of alpha BE-1, alpha BE-2, alpha BE-3, and MRF control elements. *Mol Cell Biol* **13,** 7144-7152.

Hafen E., Basler K., Edstroem J. E., and Rubin G. M. (1987) Sevenless, a cell-specific homeotic gene of Drosophila, encodes a putative transmembrane receptor with a tyrosine kinase domain. *Science* **236,** 55-63.

Haffter P., Granato M., Brand M., Mullins M. C., Hammerschmidt M., Kane D. A., Odenthal J., van Eeden F. J., Jiang Y. J., Heisenberg C. P., Kelsh R. N., Furutani-Seiki M., Vogelsang E., Beuchle D., Schach U., Fabian C., and Nusslein-Volhard C. (1996) The identification of genes with unique and essential functions in the development of the zebrafish, Danio rerio. *Development* **123,** 1-36.

Hall S. L. and Padgett R. A. (1994) Conserved sequences in a class of rare eukaryotic nuclear introns with non-consensus splice sites. *J Mol Biol* **239,** 357-365.

Halpern M. E., Ho R. K., Walker C., and Kimmel C. B. (1993) Induction of muscle pioneers and floor plate is distinguished by the zebrafish no tail mutation. *Cell* **75,** 99-111.

Harris, W. A., Stark, W. S., and Walker, J. A. Genetic dissection of the photoreceptor system in the compound eye of *Drosophila*. J.Physiol. 256, 415-439. 1976.

Heon E., Priston M., Schorderet D. F., Billingsley G. D., Girard P. O., Lubsen N., and Munier F. L. (1999) The gamma-crystallins and human cataracts: a puzzle made clearer. *Am J Hum Genet* **65,** 1261-1267.

Hirata J., Nakagoshi H., Nabeshima Y., and Matsuzaki F. (1995) Asymmetric segregation of the homeodomain protein Prospero during Drosophila development. *Nature* **377,** 627-630.

Hounsfield G. N. (1973) Computerized transverse axial scanning (tomography). 1. Description of system. *Br J Radiol* **46,** 1016-1022.

Huang S., Chen C. S., and Ingber D. E. (1998) Control of cyclin D1, p27(Kip1), and cell cycle progression in human capillary endothelial cells by cell shape and cytoskeletal tension. *Mol Biol Cell* **9,** 3179-3193.

Hudziak R. M., Barofsky E., Barofsky D. F., Weller D. L., Huang S. B., and Weller D. D. (1996) Resistance of morpholino phosphorodiamidate oligomers to enzymatic degradation. *Antisense Nucleic Acid Drug Dev* **6,** 267-272.

Ikeshima-Kataoka H., Skeath J. B., Nabeshima Y., Doe C. Q., and Matsuzaki F. (1997) Miranda directs Prospero to a daughter cell during Drosophila asymmetric divisions. *Nature* **390,** 625-629.

Irigoyen J. P., Besser D., and Nagamine Y. (1997) Cytoskeleton reorganization induces the urokinase-type plasminogen activator gene via the Ras/extracellular signal-regulated kinase (ERK) signaling pathway. *J Biol Chem* **272,** 1904-1909.

Iwaki A., Nagano T., Nakagawa M., Iwaki T., and Fukumaki Y. (1997) Identification and characterization of the gene encoding a new member of the alpha-crystallin/small hsp family, closely linked to the alphaB- crystallin gene in a head-to-head manner. *Genomics* **45,** 386-394.

Jackson I. J. (1991) A reappraisal of non-consensus mRNA splice sites. *Nucleic Acids Res* **19,** 3795-3798.

Jaynes J. B. and O'Farrell P. H. (1991) Active repression of transcription by the engrailed homeodomain protein. *EMBO J* **10,** 1427-1433.

Jowett T. (1998) Analysis of Protein and Gene Expression, in *The Zebrafish : Biology (Methods in Cell Biology, Vol 59)* (Detrich H. W., Zon L. I., and Westerfield M., eds.), pp. 63-85. Academic Press.

Kak A. C. and Slaney M. (1988) *Principles of Computerised Tomographic Imaging*, IEEE Press.

Karlovich C. A., Bonfini L., McCollam L., Rogge R. D., Daga A., Czech M. P., and Banerjee U. (1995) In vivo functional analysis of the Ras exchange factor son of sevenless. *Science* **268,** 576-579.

Karlstrom R. O., Talbot W. S., and Schier A. F. (1999) Comparative synteny cloning of zebrafish you-too: mutations in the Hedgehog target gli2 affect ventral forebrain patterning. *Genes Dev* **13,** 388-393.

Kauffmann R. C., Li S., Gallagher P. A., Zhang J., and Carthew R. W. (1996) Ras1 signaling and transcriptional competence in the R7 cell of Drosophila. *Genes Dev* **10,** 2167-2178.

Kaufman M. H. (1992) *Atlas of Mouse Development*, Academic Press.

Krauss S., Concordet J. P., and Ingham P. W. (1993) A functionally conserved homolog of the Drosophila segment polarity gene hh is expressed in tissues with polarizing activity in zebrafish embryos. *Cell* **75,** 1431-1444.

Kraut R., Chia W., Jan L. Y., Jan Y. N., and Knoblich J. A. (1996) Role of inscuteable in orienting asymmetric cell divisions in Drosophila. *Nature* **383,** 50-55.

Lai Z. C., Harrison S. D., Karim F., Li Y., and Rubin G. M. (1996) Loss of tramtrack gene activity results in ectopic R7 cell formation, even in a sina mutant background. *Proc Natl Acad Sci U S A* **93,** 5025-5030.

Lauterbur P. C. (1973) Image Formation by Induced Local Interactions: Example Employing Nuclear Magnetic Resonance. *Nature* **242**.

Lengler J., Krausz E., Tomarev S., Prescott A., Quinlan R. A., and Graw J. (2001) Antagonistic action of Six3 and Prox1 at the gamma-crystallin promoter. *Nucleic Acids Res* **29,** 515-526.

Li L. and Vaessin H. (2000) Pan-neural prospero terminates cell proliferation during drosophila neurogenesis [In Process Citation]. *Genes Dev* **14,** 147-151.

Li P., Yang X., Wasser M., Cai Y., and Chia W. (1997) Inscuteable and Staufen mediate asymmetric localization and segregation of prospero RNA during Drosophila neuroblast cell divisions. *Cell* **90,** 437-447.

Liang Z. and Biggin M. D. (1998) Eve and ftz regulate a wide array of genes in blastoderm embryos: the selector homeoproteins directly or indirectly regulate most genes in Drosophila. *Development* **125,** 4471-4482.

Link B. A., Kainz P. M., Ryou T., and Dowling J. E. (2001) The perplexed and confused mutations affect distinct stages during the transition from proliferating to post-mitotic cells within the zebrafish retina. *Dev Biol* **236,** 436-453.

Livneh E., Glazer L., Segal D., Schlessinger J., and Shilo B. Z. (1985) The Drosophila EGF receptor gene homolog: conservation of both hormone binding and kinase domains. *Cell* **40,** 599-607.

Manzanares M., Wada H., Itasaki N., Trainor P. A., Krumlauf R., and Holland P. W. (2000) Conservation and elaboration of Hox gene regulation during evolution of the vertebrate head. *Nature* **408,** 854-857.

Matsuzaki F., Ohshiro T., Ikeshima-Kataoka H., and Izumi H. (1998) miranda localizes staufen and prospero asymmetrically in mitotic neuroblasts and epithelial cells in early Drosophila embryogenesis. *Development* **125,** 4089-4098.

Mazet F. and Shimeld M. (2002) Gene duplication and divergence in the early evolution of vertebrates. *Curr Opin Genet Dev* **12,** 393-396.

Meyer A. and Schartl M. (1999) Gene and genome duplications in vertebrates: the one-to-four (-to-eight in fish) rule and the evolution of novel gene functions [In Process Citation]. *Curr Opin Cell Biol* **11,** 699-704.

Miller J. B. and Stockdale F. E. (1986) Developmental origins of skeletal muscle fibers: clonal analysis of myogenic cell lineages based on expression of fast and slow myosin heavy chains. *Proc Natl Acad Sci U S A* **83,** 3860-3864.

Nasevicius A. and Ekker S. C. (2000) Effective targeted gene 'knockdown' in zebrafish. *Nat Genet 2000* **26,** 216-220.

Neyt C., Jagla K., Thisse C., Thisse B., Haines L., and Currie P. D. (2000) Evolutionary origins of vertebrate appendicular muscle. *Nature* **408,** 82-86.

Nishiguchi S., Wood H., Kondoh H., Lovell-Badge R., and Episkopou V. (1998) Sox1 directly regulates the gamma-crystallin genes and is essential for lens development in mice. *Genes Dev* **12,** 776-781.

Norris W., Neyt C., Ingham P. W., and Currie P. D. (2000) Slow muscle induction by Hedgehog signalling in vitro. *J Cell Sci* **113,** 2695-2703.

Oliver G., Sosa-Pineda B., Geisendorf S., Spana E. P., Doe C. Q., and Gruss P. (1993) Prox 1, a prospero-related homeobox gene expressed during mouse development. *Mech Dev* **44,** 3-16.

O'Neill E. M., Rebay I., Tjian R., and Rubin G. M. (1994) The activities of two Ets-related transcription factors required for Drosophila eye development are modulated by the Ras/MAPK pathway. *Cell* **78,** 137-147.

Otake L. R., Scamborova P., Hashimoto C., and Steitz J. A. (2002) The divergent U12-type spliceosome is required for pre-mRNA splicing and is essential for development in Drosophila. *Mol Cell* **9,** 439-446.

Patel A. A., McCarthy M., and Steitz J. A. (2002) The splicing of U12-type introns can be a rate-limiting step in gene expression. *EMBO J* **21,** 3804-3815.

Piperno G. and Fuller M. T. (1985) Monoclonal antibodies specific for an acetylated form of alpha-tubulin recognize the antigen in cilia and flagella from a variety of organisms. *J Cell Biol* **101,** 2085-2094.

Postlethwait J. H., Yan Y. L., Gates M. A., Horne S., Amores A., Brownlie A., Donovan A., Egan E. S., Force A., Gong Z., Goutel C., Fritz A., Kelsh R., Knapik E., Liao E., Paw B., Ransom D., Singer A., Thomson M., Abduljabbar T. S., Yelick P., Beier D., Joly J. S., Larhammar D., and Rosa F. (1998) Vertebrate genome evolution and the zebrafish gene. *Nat Genet* **18,** 345-349.

Prince V. E., Joly L., Ekker M., and Ho R. K. (1998) Zebrafish hox genes: genomic organization and modified colinear expression patterns in the trunk. *Development* **125,** 407-420.

Raamsdonk W., van d. S., Diegenbach P. C., van d. B., Bruyn H., van D. J., and Mijzen P. (1974) Differentiation of the musculature of the teleost Brachydanio rerio. I. Myotome shape and movements in the embryo. *Anat Embryol (Berl )* **145,** 321-342.

Radon J. (1917) Über die Bestimmung von Funktionen durch ihre Integralwerte längs gewisser Mannigfaltigkeiten. *Ber Verh Sächs Akad Wiss Leipzig, Math-Nat* **69,** 262-277.

Reinke R. and Zipursky S. L. (1988) Cell-cell interaction in the Drosophila retina: the bride of sevenless gene is required in photoreceptor cell R8 for R7 cell development. *Cell* **55,** 321-330.

Rescan P. Y., Collet B., Ralliere C., Cauty C., Delalande J. M., Goldspink G., and Fauconneau B. (2001) Red and white muscle development in the trout (Oncorhynchus mykiss) as shown by in situ hybridisation of fast and slow myosin heavy chain transcripts. *J Exp Biol* **204,** 2097-2101.

Roy S., Wolff C., and Ingham P. W. (2001) The u-boot mutation identifies a Hedgehog-regulated myogenic switch for fiber-type diversification in the zebrafish embryo. *Genes Dev* **15,** 1563-1576.

Runkle S., Hill J., Kantorow M., Horwitz J., and Posner M. (2002) Sequence and spatial expression of zebrafish (Danio rerio) alphaA-crystallin. *Mol Vis* **8,** 45-50.

Schauerte H. E., van E. F., Fricke C., Odenthal J., Strahle U., and Haffter P. (1998) Sonic hedgehog is not required for the induction of medial floor plate cells in the zebrafish. *Development* **125,** 2983-2993.

Schroeder W. E., Martin K. M., and Lorenson W. E. (1997) *The Visualiszation Toolkit*, Pearson Education.

Sharpe J., Ahlgren U., Perry P., Hill B., Ross A., Hecksher-Sorensen J., Baldock R., and Davidson D. (2002) Optical projection tomography as a tool for 3D microscopy and gene expression studies. *Science* **296,** 541-545.

Shen C. P., Jan L. Y., and Jan Y. N. (1997) Miranda is required for the asymmetric localization of Prospero during mitosis in Drosophila. *Cell* **90,** 449-458.

Shen C. P., Knoblich J. A., Chan Y. M., Jiang M. M., Jan L. Y., and Jan Y. N. (1998) Miranda as a multidomain adapter linking apically localized Inscuteable and basally localized Staufen and Prospero during asymmetric cell division in Drosophila. *Genes Dev* **12,** 1837-1846.

Sosa-Pineda B., Wigle J. T., and Oliver G. (2000) Hepatocyte migration during liver development requires Prox1. *Nat Genet* **25,** 254-255.

Spana E. P. and Doe C. Q. (1995) The prospero transcription factor is asymmetrically localized to the cell cortex during neuroblast mitosis in Drosophila. *Development* **121,** 3187-3195.

Srinivasan S., Peng C. Y., Nair S., Skeath J. B., Spana E. P., and Doe C. Q. (1998) Biochemical analysis of ++Prospero protein during asymmetric cell division: cortical Prospero is highly phosphorylated relative to nuclear Prospero. *Dev Biol* **204,** 478-487.

St.Johnston D., Driever W., Berleth T., Richstein S., and Nusslein-Volhard C. (1989) Multiple steps in the localization of bicoid RNA to the anterior pole of the Drosophila oocyte. *Development Suppl* **107,** 13-19.

Stickney H. L., Barresi M. J., and Devoto S. H. (2000) Somite development in zebrafish. *Dev Dyn* **219,** 287-303.

Streicher J., Donat M. A., Strauss B., Sporle R., Schughart K., and Muller G. B. (2000) Computer-based three-dimensional visualization of developmental gene expression. *Nat Genet* **25,** 147-152.

Summerton J. and Weller D. (1997) Morpholino antisense oligomers: design, preparation, and properties. *Antisense Nucleic Acid Drug Dev* **7,** 187-195.

Talbot W. S., Trevarrow B., Halpern M. E., Melby A. E., Farr G., Postlethwait J. H., Jowett T., Kimmel C. B., and Kimelman D. (1995) A homeobox gene essential for zebrafish notochord development. *Nature* **378,** 150-157.

Tomarev S. I., Sundin O., Banerjee-Basu S., Duncan M. K., Yang J. M., and Piatigorsky J. (1996) Chicken homeobox gene Prox 1 related to Drosophila prospero is expressed in the developing lens and retina. *Dev Dyn* **206,** 354-367.

Tomarev S. I., Zinovieva R. D., Chang B., and Hawes N. L. (1998) Characterization of the mouse Prox1 gene. *Biochem Biophys Res Commun* **248,** 684-689.

Trower M. K., Orton S. M., Purvis I. J., Sanseau P., Riley J., Christodoulou C., Burt D., See C. G., Elgar G., Sherrington R., Rogaev E. I., George-Hyslop P., Brenner S., and Dykes C. W. (1996) Conservation of synteny between the genome of the pufferfish (Fugu rubripes) and the region on human chromosome 14 (14q24.3) associated with familial Alzheimer disease (AD3 locus). *Proc Natl Acad Sci U S A* **93,** 1366-1369.

Tsuda L., Inoue Y. H., Yoo M. A., Mizuno M., Hata M., Lim Y. M., Adachi-Yamada T., Ryo H., Masamune Y., and Nishida Y. (1993) A protein kinase similar to MAP kinase activator acts downstream of the raf kinase in Drosophila. *Cell* **72,** 407-414.

Vaessin H., Grell E., Wolff E., Bier E., Jan L. Y., and Jan Y. N. (1991) prospero is expressed in neuronal precursors and encodes a nuclear protein that is involved in the control of axonal outgrowth in Drosophila. *Cell* **67,** 941-953.

van den Heuvel M. and Ingham P. W. (1996) smoothened encodes a receptor-like serpentine protein required for hedgehog signalling. *Nature* **382,** 547-551.

van Eeden F. J., Granato M., Schach U., Brand M., Furutani-Seiki M., Haffter P., Hammerschmidt M., Heisenberg C. P., Jiang Y. J., Kane D. A., Kelsh R. N., Mullins M. C., Odenthal J., Warga R. M., Allende M. L., Weinberg E. S., and Nusslein-Volhard C. (1996) Mutations affecting somite formation and patterning in the zebrafish, Danio rerio. *Development* **123,** 153-164.

van Raamsdonk W., Pool C. W., and te Kronnie G. (1978) Differentiation of muscle fiber types in the teleost Brachydanio rerio. *Anat Embryol (Berl )* **153,** 137-155.

Varga Z. M., Amores A., Lewis K. E., Yan Y. L., Postlethwait J. H., Eisen J. S., and Westerfield M. (2001) Zebrafish smoothened functions in ventral neural tube specification and axon tract formation. *Development* **128,** 3497-3509.

Wada H., Garcia-Fernandez J., and Holland P. W. (1999) Colinear and segmental expression of amphioxus Hox genes. *Dev Biol* **213,** 131-141.

Wallace K., Liu T. H., and Vaessin H. (2000) The pan-neural bHLH proteins DEADPAN and ASENSE regulate mitotic activity and cdk inhibitor dacapo expression in the Drosophila larval optic lobes. *Genesis* **26,** 77-85.

Wasiak S., Quinn C. C., Ritter B., de Heuvel E., Baranes D., Plomann M., and McPherson P. S. (2001) The Ras/Rac guanine nucleotide exchange factor mammalian Son-of- sevenless interacts with PACSIN 1/syndapin I, a regulator of endocytosis and the actin cytoskeleton. *J Biol Chem* **276,** 26622-26628.

Waterman R. E. (1969) Development of the lateral musculature in the teleost, Brachydanio rerio: a fine structural study. *Am J Anat* **125,** 457-493.

Weinberg E. S., Allende M. L., Kelly C. S., Abdelhamid A., Murakami T., Andermann P., Doerre O. G., Grunwald D. J., and Riggleman B. (1996) Developmental regulation of zebrafish MyoD in wild-type, no tail and spadetail embryos. *Development* **122,** 271-280.

Weninger W. J. and Mohun T. (2002) Phenotyping transgenic embryos: a rapid 3-D screening method based on episcopic fluorescence image capturing. *Nat Genet* **30,** 59-65.

Wigle J. T., Chowdhury K., Gruss P., and Oliver G. (1999) Prox1 function is crucial for mouse lens-fibre elongation. *Nat Genet* **21,** 318-322.

Wigle J. T. and Oliver G. (1999) Prox1 function is required for the development of the murine lymphatic system. *Cell* **98,** 769-778.

Williams B. A. and Ordahl C. P. (1994) Pax-3 expression in segmental mesoderm marks early stages in myogenic cell specification. *Development* **120,** 785-796.

Wilson S. W. and Easter S. S., Jr. (1991) Stereotyped pathway selection by growth cones of early epiphysial neurons in the embryonic zebrafish. *Development* **112,** 723-746.

Xu Y., He J., Tian H. L., Chan C. H., Liao J., Yan T., Lam T. J., and Gong Z. (1999) Fast skeletal muscle-specific expression of a zebrafish myosin light chain 2 gene and characterization of its promoter by direct injection into skeletal muscle. *DNA Cell Biol* **18,** 85-95.

Xu C., Kauffmann R. C., Zhang J., Kladny S., and Carthew R. W. (2000) Overlapping activators and repressors delimit transcriptional response to receptor tyrosine kinase signals in the Drosophila eye. *Cell* **103,** 87-97.

Zinovieva R. D., Duncan M. K., Johnson T. R., Torres R., Polymeropoulos M. H., and Tomarev S. I. (1996) Structure and chromosomal localization of the human homeobox gene Prox 1. *Genomics* **35,** 517-522.

# Appendices

## Appendix 1 Sequence

### 1.1 Primer sequences

| Name | Sequence | Purpose |
|---|---|---|
| *prox1* 5'F | CTGACCATGACAGCACATCC | Amplify first 400 bp |
| *prox1* 5'R | AATGGTGAAAGGCACTCCTG | |
| *prox1* 5' SalI | GAAGATGGGTCGACAACAGTTCCAG | Clone *prox1* with 5' SalI site |
| *prox1* 3' ClaI | ATGGGGTAAATCGATAATACAGTC | Clone *prox1* with 3' ClaI site |
| *prox1* 3' Xba1 | ATGGGGTAATCTAGAAATACAGTC | Clone *prox1* with 3' XbaI site |
| *prox1* 5' EcoR1 | GAAGATGGGAATTCAACAGTTCCAG | Clone *prox1* with 5' EcoRI site |
| *prox1* 3' SacII | ATGGGGTAACCGCGGAATACAGTC | Clone *prox1* with 3' SacII site |
| *prox1* probe mid F | GACCCAATTTGAGATGGAGC | Amplify bp 447-883 |
| *prox1* probe mid R | GATCTGGGACAGAGTCGTGG | |
| *prox2* T7-2 primer | GCTGAACCCATCTGTTCACA | Sequencing primer |
| *prox2* T3-2 primer | CCTCTCATGCATCCCAATAAC | Sequencing primer |
| *prox2* T3-3 primer | TTCTCCTCAACGTTGTTCC | Sequencing primer |
| *prox2* T7-5 primer | GGATATAGAAGAACTCCCGG | Sequencing primer |
| *prox2* reverse primer | CCATTCTTAGTCTCTTAAC | Sequencing primer |
| *prox2* start primer | GCAATAACAGGAAACTCAGTGG | Sequencing primer |
| 5' *prox3* internal primer | GCGTGGAGAACATCATACGG | Sequencing primer |
| 3' *prox3* internal primer | CCTGTCATCATCAAGTTCTTCG | Sequencing primer |
| *Fugu* 5' *prox4* | CAGAAGCCTGCTGGTGGATCC | Amplify fragment of *Fugu prox4* to screen library |
| *Fugu* 3' *prox4* | TTCTCTCAGTGTGATCGCAGC | |
| *Fugu* 5' *prox4-2* | GTGAGCTTTCCAGAGCCATC | |
| *Fugu* 3' *prox4-2* | TTCTTCAGATGGTTGGTGGTC | |
| 5' *Fugu prox4* internal primer | GGAGCATCGACTCCATATTCA | Nested primers for product of *Fugu prox4* amplification |
| 3' *Fugu prox4* internal primer | CGTTAAGGACTAAACGGGGG | |
| *prox4* conserved region 3' primer1 | GAACCACTTAATCAGCTGG | Primers against conserved *Fugu* and *Tetraodon prox4* sequence to amplify from zebrafish |
| *prox4* conserved region 3' primer2 | TTGTAGTGCATGTTGAGA | |

## 1.2 Sequences

### 1.2.1 Zebrafish *prox2*

>ZebrafishProx2
MDSPSDRFHQQPSQMCPFELYRSLPDSNPSGHRFPPGSLSSHPLFSPLMHPNSMSQKWGSFRQRQGLF
PGLLVDDDVSEEDVFGKREFGATSSSQRGSGERGQAPEPSDWGQDFARVKRLRMDSAARSEDESIREG
KRGRKSRSSFRMDEKKGMEQGNRRESREGRKRQRQELKLQLEETRGKLLELQRKVWRVYGGQADDEKD
QENGGINLDEGADMFSDSDDSLVGNGFSPQRCSSKKNNETNGNGNAGIFPETSVDLDLELNGSQVWLG
CNLIRGEWESAKGGQKFAQALKQELANVVAQVIDRVVRLYAESDPEAVSSAQEPSMTMDLNSDRRPRP
QAVEQVEALPLVTKSPQDKRAPVAQSGHKDLLLPQANPSLAPLAQPPLPALLPPRGKEHFLPSYPPDN
PVHLPLLHYTMQNLFARSLSSLPLHKDCLSESFMDFRSHNFAFPPLPLLGQLDPSPSDRAREVGMRAG
GAGMVDGADAALYLAAGSSQEGLSPCHLKKAKLMFFYTRYPSSSTLKTYFPDVKFNRCVTSQLIKWFS
NFREFFYIQMERFARQAAREVLTSARERALRLGRDTELYRILNMHYNKSNDYQVPERFVEISEVALRE
FFTAIQSGRDADPCWKKSIYKIICKLDSPVPDSFRLPGCPTDTYRMG

>Zebrafish*prox2*
atggattccccttcggatcgatttcatcagcagccttcccagatgtgtcccttcgagctgtatcgcag
cctgccagattccaacccctcaggccaccgatttcccccccggcagcttgtcttctcatccgctcttct
cccctctcatgcaccccaatagcatgtcccaaaagtggggcagtttccgacagcgtcagggcttgttc
cctgggcttctggtggatgatgatgtaagtgaggaggatgtctttgggaagagggagtttggggcgac
ctcatcatcacagcgtggttctggagaacggggacaagctcctgagccctctgactgggggcaggatt
ttgcaagagttaagagactaagaatggactctgcagctagaagcgaagatgaatcaattagagaaggg
aagagagggagaaagagtagatcttcatttaggatggatgagaagaaaggaatggagcaaggtaacag
gagggaaagcagggaaggaaggaagcggcagcggcaagaactaaagctccagctggaggaaaccagag
ggaaacttcttgaacttcarcgcaaggtgtggagggtgtacggagggcaagcggacgatgagaaggat
caagaaaacggaggaatcaatctggatgagggtgctgatatgttttctgacagcgacgattcacttgt
aggcaatggattttctcctcaacgttgttcgtctaagaaaaacaatgaaaccaatgggaatggcaatg
ctggaatttttccagagacctccgtggacctggacctggagttaaacggcagtcaagtgtggctggga
tgcaatttgatcaggggcgagtgggagagcgcagagggagggcagaagtttgctcaagcgctgaagca
ggagcttgccaatgttgtcgctcaggttattgacagagttgttcgtctctatgctgaaagtgaccccg
aggcagtctcatccgcccaggaacccagcatgacgatggatctaaactcagacaggaggccgagacct
caagctgtagagcaggtagaagctttgccactcgtcacaaagagtcctcaagacaagagggcccctgt
cgctcaaagtgggcacaaggacctcctcctccctcaagccaatcctagcctcgctccgctagcccagc
cgcctctcccggcacttttaccccccaggggcaaggagcatttcctgccctcttatcccccagacaac
cttgtccacttgccccttcttcactacaccatgcagaatctctttgcccgctccctctccagtctgcc
tcttcataaggactgcctgtctgaatctttcatggacttccgttcccacaactttgcgttcccacccc
tcccgctgctgggccagctcgacccctctccatcagatagagcaagggaagtgggcatgagggcaggt
ggtgcgggaatggtggatggtgcagatgcagcactctatcttgctgcaggttcatctcaggaggccct
ttctccgtgccacttgaggaaagccaagctcatgtttttctacacgcgttaccccagctccagcacac
tgaagacatatttccctgacgtcaagttcaatcgctgtgtcacttcccagctgattaagtggttcagt
aacttccgggagttcttctatatccagatggagcgctttgcccggcaggcagccagagaggtgcttac
atctgccagagaaagggctctccgcctgggcagagacaccgagctttaccgcattctcaacatgcact
acaataagagcaatgattaccaggttccagagaggtttgtggagatatctgaagtggccctgagggag
tttttcaccgccatccagagcggtcgggatgctgatccatgctggaagaagtccatctataagatcat
ctgtaaattagacagtccagtgcccgacagcttccggctgcccggatgccccacggatacttacagga
tgggttaa

### 1.2.2 Zebrafish *prox3*

>Zebrafish*prox3*
GGTGTCGTCCTGCAGTACGGTCCGGAATTCGCGGGTCGTCCCACGCGTCCGCAAAGATCAGCAGTTTA
CGGGCTCATATTCACCTGTAAAATCATGAATCTGAGTCCACCTGAAGCGAACATGCACAAAGTGAGCG
ACGGCGGCTGCATGGAGGTCAGCAAACCCGATCATCTGCTGCCGTGCTTTCGCAGAAGCGCATACGAA
GAGCCTCTCGCTTCCTACCACAGCGGCTCCATCATCTCTCACCTCCTGCGCAAAACCATCCACAACAA
GCTGACGTCGCTGGACAGCAGCCTCCTGTACCTTCCGTCTACTAACGCTGACTCCGAGTCCAACACGA
CGGATGATTTCCCAGCAGAGGACCAGAGCGGCAGTGAGGATCTGTCTGTGGGTGTGGACGCTGAGCGG
CCTCTAAATGAACACCTCCAGGCCAAACGCGCCCGCGTGGAGAACATCATACGGNGAATGGCAGGTTC
ACCAAACGCCAGGACACACGGGGAAGCAGATAGCGACTGTGGAAAGGAGAACAAGCGGGAGCAAAGGC
TCCCGCAGCATCAGGAACCCAGCAGGAACCAATCCAGCAAAGACGAAGAGTGCCACAAGCTGAAAGAG
CAGCTCCAGAGCATGCAACGCCTGCTGCATCAGCTCCAGGAGGAAAACCCCGCCCGTCTACCGATCAA
AACTATCTGGGGAACTAAAAACCTGGTGGTGCAAGGAGGCGAAACACCAACCTTTTTGTTTTACACTC
GGAAGAAGCGGGGGAAAAAGTCCCCCCGCAACACC

>Zebrafish*prox3*
GACCTTTTACAGGTTATTTCCAAATTGACATTGCGTTTCCGAGGAGTTCAGTTTTCCCTCCACCTGTA
AAGAGAGAGAGAGAGCCCACCTTCAGGACACTTTCATCCACTCCTTCATCACACATGCTCTTTTTATA
GGTCATTAATCAGCTTTTATTCAAATCAACATGTGTATCTGCTGTGTGTGTTGTTTCACATCGTCAGT
GGGATGGGCTTGTCTATTAATAGTTTGCAGTTTTGGTTTCAGTGTAAGATATTTTCTGCTTTGTAAAC
CACCGTGGACGAATGTGTGGTGTGTTTTCAGGTCAACAACGCTCCTCGTTATGCAGATTTCGTGGTTC
AATGTGGATTTGTCTGTTTACTTTCTGTTTTCGAAATCTAATGACACACAATCATGATCTGTAAGATA
TTTAGAAGGAAACGTGTAAATAAGTATACTTTAGCTGTATGATGAACACTGCCGTTTACATCAGATGA
CTATATATGATTGCATGTAGAAGCATTTCCCTCAAATCACTCTTATTTATCCTGTATAGTTTGCATAT
TTTTGTTACTCATTATCTAGTGGTGTGTGGCTTTTCATTTCACAAGGGTTCACATTGTAGCAGTTTGT
TTTGTTGCCGTTGAGGATTCTGATATAATTTATTTCTTTAATTTGGAAGAAATGTAAAACAATAAATA
AAGAACCAACCAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA

### 1.2.3 Zebrafish Slow Myosin Heavy Chain

Incorporating sequencing of sMyHC ESTs fc93a05 and fc92b07, carried out by D.Keenan, and EST sequences in the same cluster as identified by Washington University Zebrafish Genome Resources.

>Zebrafish-sMyHC-Contig1
CGAACATTATTAAGACATCCTATAGATCATCTATGACGAACCATGCACGCGTACGTAAGCTTGGATCC
TCTAGATCGGCCGCCCTTTTTTTTTTTTTTTTTGGTAATACAATATTTTCCTTTCAATACGAACACAAT
TGAAAAAACTCAGACTTTTTTTTTTTTTACATAATGCTATGCATCTTTGGCTCAGCCCTGACCATGCA
AGAGGCTGTAGCATTTTGCCTCAAGCGATATTCTTTAACTGAATCCTGACTGGAGTAAGAATGCAAGC
ACAACGTTTGCTAAAGTCACTGCTTATTTAGAAACAAAGAGGAAGAATGTCAGAAGTAAAAGCACACA
AACAGGAAAGAGGAATCAACAGTGCTGTGAATCATGATGCTGTACAAACCCTTTAAAGCACATGTGAA
ATCATAATATTAAAAACCATTCTCCACCCCCTCCCTCGCTGGATCCCGGATTGTGGAAGACGTTGATA
TTATACTCAAAAGTCTTCCTTCCTGTGAACGTGAGTGCATGTGAGTGTGCGCTATGCGGTGGTCTCGC
CATCGCTGTCGTTTAGGTAGCTGCTGCGTTTGTAGCGTGAGCGGTAGGTGCTTCGGCTCGTGCTCCCT
TCCTCTTTTTCTTCTTCCTCTTCCTCGNCAGATCTGTCAAGGCAAAAGCGTCTGCGGCTGGAGGGTGG
GCTTGAGCTCTTAAACTACCATCATCTGACAGATTCTTGGCAGAGCCCTTGTTTTTGGACAGCCATGA
CTTCACCCCATCAACACGGTCCTTCACCTCAGAGTCCGTATCAGAGGCA

>Zebrafish-sMyHC-Contig2
CCGGAATTACCGGGTCGACCCACGCGTCCGGAGAAAGCTGGAGGATGAATGCTCTGAACTCAAGAAAG
ACATTGATGATCTGGAGCTCACTCTGGCCAAAGTGGAGAAGGAGAAACATGCAACTGAGAACAAGGTT
AAGAACCTGACTGAAGAGATGGCAGCTTTGGATGAGATTATCGCTAAGCTGACCAAGGAGAAGAAAGC
TCTACAGGAGGCCCATCAGCAAACACTGGACGACCTCCAGAGTGAGGAAGACAAAGTCAACACACTCA
CCAAAGCCAAAGCCAAGCTGGAGCAGCAAGTGGATGATCTTGAAGGATCGTTGGAGCAAGAAAAGAAG
CTCCGCATGGATCTTGAAAGGGCAAAGAGAAAGCTGGAGGGTGATTTAAAATTGACCCAGGAGAGCAT
TATGGACTTGGAAAATGACAAACAGCAAATGGAGGAGAAGCTAAAGAAGAAGGACTTTGAAATTAGCC
AGCTCAACAGCAAAATTGAGGATGAGCAAGCTCTTGGTGCCCAGCTCCAGAAGAAACTGAAGGAGCTT
CAGGCCCGAATTGAAGAGCTTGAGGAAGAGCTGGAGGCTGAAAGAGCTGCTCGTGCCAAAGTTGAGAA
ACAGAGAGCCGATCTGTCCAGAGAACTGGAGGAGATCAGCGAGAGGTTGGAGGAGGCTGGTGGAGCCA
CCGCTGCCCAGATCGAGATGAACAAGAAACGTGAAGCTGAGTTCCAGAAACTGCGCAGAGACCTTGAA
GAGGCCACTCTGCAGCATGAGGCCACTGCTGCAACACTGAGGAAGAAACATGCCGACAGTGTGGCTGA
TCTGGGAGAGCAGATCGACAACCTTCAGAGAGTCAAGCAGAAGCTGGAGAAAGAGAAGAGCGAGCTGA
AGTTAGAGCTTGATGATGTGGTCTCCAACATGGTTCAGATTGTCAAGTCAAAGAGCAATCTGGAGAAA
ATGTGCAGAACTCTGGAGGACCAGATGAGTGAGTACAGAACCAAAGCAGAGGAAGGACAGCGCACAAT
CAACGATTTCACCATGCAAAAAGCCAAGCTGCAAACTGAGAATGGTGAACTGTCCAGACAGCTGGAGG
AGAAAGACTCCCTGGTGTCTCAGTTGACCAGAGGAAAGCAGTCCTACACTCAGCAGATTGAAGACCTC
AAGAGACAGCTAGAGGAGGAAGTCAAGGCAAAGAATGCCCTGGCACATGCAGTTCAGTCTGCTCGTCA
TGATTCTGACCTGCTGAGGGAGCAGTTTGAGGAGGAGCAGGAAGCCAAAGCTGAGCTGCAGCGTAGTC
TGTCCAAGACAAACTCTGAAGTGGCTCAGTGGAGAACCAAGTATGAAACTGATGCCATCCAGAGGACT
GAGGAGCTGGAGGATGCCAAGAAAAATTGGCACAGCGTCTCCAGGAAGCAGAAGAGGCTGTGGAAGCT
GNTNAATGCTAATGCTCCTCTCTGGAGAAGACCAGCACAG

>Zebrafish-sMyHC-Contig3
CCCAGCTTCAAGCCTAAGGTTAAGGAGCCCGTGCAAGAAGTCAAGGAATGTTGAGAAAAGGCCCAAGA
AAGCCCATCACTAATGCTGCCATAATGCCAGAGAAGCTGAGAAAGGAGCAGAACACCAGTGCTCATCT
GGAGCGCATGAAGAAGAACATGGAGCAGACCATCAAGGACTTGCAGCACCGTCTGGATGAAGCTGAGC
AGATCGCCATGAAGGGTGGCAAGAAGCAGGTCCAGAAACTGGAAGCCAGGGTGAGAGAGCTGGAGAAT
GAGGTGGAGTTGGAACAGAGAAAGGCGAGCGAGTCTGTAAAAGGAGTCCGTAAATATGAGAGGCGCAT
CAAGGAGCTCACCTACCAGACTGAAGAAGACCGTAAGAATCTGGCTCGTCTTCAAGATCTGGTGGACA
AACTGCAGCTGAAGGTCAAGTCCTACAAGAGAGCTGCTGAAGAGGCTGAGGAACAGGCCAATTCTAAC
CTGGGCAAGTTCCGTAAGCTGCAGCATGAGCTGGACGAGGCAGAGGAGAGGGCTGATATTGCTGAATC
TCAGGTCAACAAGATGAGAGCCAAGAGCCGTGATTCAGGACCCAAGAAAGGACATGATGAAGAGTAAT
CATCCAGTATAACGAGTTTGTTTGTTGTGAAACCATGTTCAAGTACTGGTTTTAATCTAGCAATAAAA
ACTGCAATTTTGAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
AAA

## 1.3 Predicted protein and coding sequences

### 1.3.1 Human *proxB*

>HumanProxB
QTESITASCRLTQIKGCGQHAGLSTEVSCFSAAQRQSFHDCMGVNISGYPATVESGWISSFWPQDISV
HCCFIKHTGPICYPSRSVFWSTQTTRNLRNNTLNNPQGRFSSFSGSMLFSSMTLSTRLSFCLQLDRDS
PFPWSQVPSSSPTDPEWFGDEHIQAKRARVETIVRGMCLSPNPLVPGNAQAGVSPRCPKKARERKRKQ
NLPTPQGLLMPAPAWDQGNRKGGPRVREQLHLLKQQLRHLQEHILQAAKPRDTAQGPGGCGTGKGPLS
AKQGNGCGPRPWVVDGDHQQGTSKDLSGAEKHQESEKPSFLPSGAPASLEILRKELTRAVSQAVDSVL
QKVLLDPPGHLTQLGRSFQGQVAEGRSEPSPPVGGACKDPLALAALPRRVQLQAGVPVGNLSLAKRLD
SPRYPIPPRMTPKPCQDPPANFPLTAPSHIQENQILSQLLGHRYNNGHWSSSPPQDSSSQRHPSSEPA
LRPWRTTKPQPLVLSQQQCPLPFTSAHLESLPLLPSVKMEQRGLHAVMEALPFSLVHEGLNPGHLKKA
KLMFFFTRYPSSNLLKVYFPDVQFNRCITSQMIKWFSNFREFYYIQMEKSARQAISDGVTNPKMLVVL
RNSELFQALNMHYNKGNDFEVPDCFLEIASLTLQEFFRAVSAGRDSDPSWKKPIYKIISKLDSDIPEI
FKSSSYPQ

>Human*ProxB*
CAAACAGAAAGCATCACTGCTTCATGCAGACTTACCCAGATAAAGGGCTGTGGGCAGCACGCCGGCCT
CAGCACAGAGGTTTCCTGTTTCTCAGCAGCCCAGAGGCAGAGTTTCCATGATTGCATGGGCGTAAACA
TCTCTGGGTACCCAGCCACAGTGGAGTCTGGCTGGATTTCCAGTTTCTGGCCTCAAGATATTTCTGTT
CATTGTTGCTTCATTAAACACACTGGTCCCATCTGTTACCCTAGTAGATCTGTCTTCTGGAGCACACA
GACTACACGGAACCTCAGAAACAATACCTTGAATAACCCACAAGGGAGATTCAGCAGCTTTAGTGGTT
CTATGCTCTTCTCCTCAATGACTCTATCTACCCGTCTGTCTTTCTGTCTGCAGCTGGATAGAGACTCC
CCGTTTCCCTGGAGTCAGGTCCCCAGCTCCAGCCCTACAGACCCCGAATGGTTTGGTGATGAGCACAT
CCAGGCAAAGAGGGCCAGAGTGGAGACCATTGTCCGAGGCATGTGTCTCTCCCCGAATCCTCTGGTGC
CAGGCAATGCGCAAGCTGGGGTCAGCCCACGCTGCCCAAAGAAGGCCCGAGAGAGGAAGAGGAAGCAG
AACCTTCCCACACCGCAAGGCCTCCTGATGCCAGCCCCTGCCTGGGACCAGGGCAACAGGAAGGGGGG
CCCTCGTGTGAGAGAACAACTTCATCTGCTGAAGCAACAGCTAAGACATCTGCAAGAGCACATCCTAC
AGGCTGCCAAGCCCAGGGACACAGCTCAGGGGCCAGGAGGCTGTGGCACGGGGAAAGGCCCTCTGAGT
GCAAAGCAGGGGAATGGCTGTGGGCCTCGCCCCTGGGTTGTGGACGGTGACCACCAGCAAGGTACCAG
CAAGGACCTCTCTGGGGCAGAAAAACACCAAGAGTCTGAGAAGCCCAGTTTCCTTCCTTCTGGAGCAC
CAGCTTCACTAGAGATTCTGAGGAAAGAGCTGACCAGGGCAGTGTCCCAGGCTGTGGACTCGGTATTA
CAAAAGGTACTATTGGATCCACCAGGCCACCTGACTCAGCTGGGCAGAAGCTTCCAGGGGCAGGTGGC
AGAGGGTAGAAGCGAGCCCTCACCTCCTGTGGGAGGGGCCTGTAAAGATCCACTTGCTTTGGCTGCCT
TGCCCAGGAGGGTCCAGCTACAAGCTGGGGTCCCAGTAGGAAATTTATCACTGGCCAAGCGTCTAGAT
TCTCCTAGGTACCCTATCCCTCCAAGAATGACCCCCAAACCCTGTCAGGATCCCCCAGCAAACTTTCC
CTTGACTGCACCTTCCCACATCCAGGAAAATCAGATTCTTAGCCAGCTACTGGGTCATAGATACAACA
ATGGCCATTGGAGTAGCAGTCCTCCCCAGGACTCATCTTCCCAGAGGCACCCCTCCTCAGAGCCTGCC
CTACGACCTTGGAGAACTACTAAGCCGCAACCATTGGTCCTGAGCCAGCAGCAGTGTCCCTTGCCTTT
CACCTCTGCCCATCTGGAAAGTCTACCCCTTCTTCCCTCGGTGAAGATGGAACAGAGAGGCCTGCATG
CTGTCATGGAGGCACTGCCTTTCTCTTTGGTCCACGAGGGTCTAAACCCTGGTCACTTGAAGAAGGCC
AAACTAATGTTTTTCTTCACACGATATCCCAGCTCCAACCTCCTGAAGGTTTATTTTCCTGATGTTCA
GTTCAACCGCTGCATTACCTCCCAGATGATCAAGTGGTTCAGCAACTTTCGTGAGTTTTATTACATCC
AAATGGAAAAATCTGCCCGGCAAGCAATTTCAGATGGTGTCACAAATCCCAAAATGCTGGTGGTTCTC
CGCAATTCAGAACTTTTTCAAGCTCTCAATATGCACTACAACAAGGGAAATGACTTTGAGGTTCCAGA
TTGCTTCTTGGAAATTGCCAGCTTGACGTTACAGGAGTTCTTCAGGGCTGTCTCCGCAGGGAGAGACT
CAGATCCTTCCTGGAAGAAACCCATTTATAAAATTATTTCGAAACTGGACAGTGACATCCCAGAGATA
TTCAAATCTTCCAGCTATCCCCAG

### 1.3.2 Mouse *proxB*

>MouseProxB
MPRADITGRPTNIRRARLRVIGSRNKSALPVK*GSITLAFLNPGAVDVHSLRVKKTLQHALAHLSLLN
LAGLSKALGSSWELCMSIRRGRTWIHLLLSCFLLSLGPAPTWQTCMDQERSPATAEAGRDSFPSGQLP
SSSLTEADWFWDEHIQAKRARVETIVRGMCLSPSSSVSGRARESLRCPEKGRERKRKQSLPMHQGPLK
SSPAWERGPKKGGTRVKEQLHLLKQQLRHLQEHVLQATEPRAPAQSPGGTEPRSSPRARPRNSCSSGA
WTVENEPHQSSSKDLCGAVKPGAAEVLQYSEEPMLCPSGPRALVETLRKELSRAVSQAVDSVLQQVLF
DPQRHLTQQERSCQGLASEGRNQPSPPGRSAYKDPLALATLPRKIQPQAGVPLGNSTLARPLDSPMCP
VSPRGVPRSYQSPLPNCPLTNVPSHTWENQMLRQLLGRGPDGQWSGSPPQDAAFQSHTSPESAQQPWG
LSQQQLPLSLTPVHLESRPLPPPVKMEQGVLRGVADSLPFSSIHEGLSPGHLKKAKLMFFFTRYPSSS
LLKAYFPDVQFNRCITSQMIKWFSNFREFYYIQMEKYARQALSDGITNAQALAVLRDSELFRVLNTHY
NKGNDFEVPDCFLEIAALTLKEFFRAVLAGKDSDPSWKKPIYKVISKLDSDVPEMLKSPSFLPGLFPS

>Mouse*ProxB*
ATGCCAAGGGCAGATATCACTGGCAGGCCTACAAACATCAGAAGAGCAAGGCTCAGGGTAATAGGAAG
CAGAAATAAGTCAGCTCTCCCAGTCAAGTAGGGTTCCATCACTCTGGCCTTCCTGAACCCCGGAGCAG
TTGATGTTCACTCTCTGCGTGTGAAGAAGACCCTTCAGCACGCTCTGGCCCACCTCTCGCTGCTCAAT
TTAGCAGGGCTCTCAAAGGCCCTGGGCTCCTCCTGGGAGCTCTGCATGAGCATCCGGAGAGGCAGGAC
ATGGATCCACCTGCTGCTGTCTTGCTTCCTCCTCAGTCTCGGACCTGCACCCACCTGGCAGACCTGCA
TGGACCAGGAGAGAAGCCCAGCCACCGCAGAGGCAGGCAGAGACTCGTTTCCCAGTGGTCAGCTGCCC
AGCTCCAGCCTCACTGAAGCTGACTGGTTTTGGGACGAGCACATCCAGGCTAAGAGGGCCAGAGTAGA
GACCATTGTCCGAGGCATGTGTCTCTCTCCTAGCTCGTCGGTGTCAGGCAGAGCCAGGGAGAGCTTGC
GCTGTCCAGAAAAGGGTCGGGAGCGGAAGAGAAAGCAGAGCCTTCCCATGCACCAAGGTCCCCTGAAA
TCTAGCCCTGCCTGGGAGCGAGGGCCCAAGAAGGGGGGGACCCGGGTGAAAGAACAGCTTCATCTACT
GAAGCAACAGCTAAGACATTTGCAGGAGCATGTCCTACAGGCCACAGAGCCCAGAGCACCAGCACAGA
GCCCAGGAGGAACTGAGCCAAGAAGCTCCCCGAGGGCAAGGCCGAGGAACAGCTGCTCATCTGGTGCC
TGGACCGTGGAGAACGAGCCTCACCAGAGTTCTAGCAAGGACCTCTGTGGAGCAGTAAAGCCCGGAGC
AGCTGAGGTCCTACAGTACTCAGAGGAGCCGATGCTCTGTCCTTCTGGCCCACGGGCCTTGGTGGAGA
CTCTGAGGAAAGAACTGAGCAGGGCCGTGTCCCAAGCTGTGGACTCCGTATTACAGCAAGTGCTATTT
GATCCACAGAGACATCTCACGCAGCAGGAGCGAAGCTGCCAGGGGCTAGCATCAGAGGGCAGAAACCA
GCCTTCACCTCCAGGGAGAAGTGCCTATAAGGACCCACTTGCTCTTGCCACCCTGCCCAGAAAGATCC
AGCCACAAGCTGGGGTTCCCTTGGGCAATTCAACACTGGCAAGGCCTCTAGATTCTCCTATGTGTCCT
GTCTCTCCAAGAGGAGTCCCTAGATCATATCAGAGTCCCCTACCAAATTGTCCCTTGACAAATGTACC
TTCCCACACCTGGGAAAATCAGATGCTTAGGCAGCTTCTGGGTCGTGGGCCCGATGGCCAATGGAGTG
GCAGCCCTCCCCAGGACGCAGCTTTCCAAAGTCATACCTCCCCAGAGTCTGCCCAGCAACCGTGGGGG
CTGAGCCAGCAGCAGCTCCCCCTGTCTCTCACCCCGGTCCATCTGGAAAGCCGGCCCCTCCCACCGCC
TGTGAAGATGGAGCAGGGCGTGCTGCGGGGCGTGGCTGACTCACTTCCTTTTTCTTCCATCCACGAAG
GCCTGAGCCCTGGTCACCTGAAGAAGGCCAAACTAATGTTTTTCTTCACGCGGTACCCCAGCTCCAGC
CTCCTGAAAGCTTATTTCCCTGATGTTCAGTTCAACCGCTGCATCACCTCCCAGATGATCAAGTGGTT
CAGCAACTTCCGTGAATTTTATTACATCCAGATGGAAAAGTATGCCCGGCAAGCACTTTCAGACGGTA
TCACGAACGCCCAAGCACTGGCTGTCCTCCGAGACTCGGAGCTTTTCCGAGTTCTCAATACACACTAT
AACAAGGGGAATGACTTTGAGGTCCCAGACTGCTTCTTGGAGATTGCCGCCCTGACACTGAAGGAGTT
CTTCAGGGCTGTCTTGGCAGGGAAAGATTCAGATCCTTCCTGGAAGAAACCCATTTACAAGGTTATCT
CAAAACTCGACAGTGACGTCCCAGAGATGCTCAAATCACCCAGCTTCCTCCCGGGACTGTTCCCGAGT

### 1.3.3 Fugu *prox2*

>FuguProx2
FAVMDSPSDLFEDSGRQIHTFASTLSSSDLTGVQLQPSTRHPPPAFRPLGFPLIHNLLHPGGATRRVG
GQLNPNLRSHRHLEEVNLEENMQERDGRVDHEMDGLIEGEDSLLVVKKRHSDTVLLGEWSQDVLKVKR
VRNGEPEEGGKKNQAGRKQKRREREELKEQLEEARERLQALQEKVWKVFGEKHRVEEERSQRGNRDAG
DEGMVEEEDMFEEDDGGDLEKENFSLLSGSPFENFHKHREEKQKDTGGRTERE REGQLHLEGATDGAD
LTMEAEDWNGMEDEGEEGGQKFAQALKLELGSAVARVIDRVLRLYTEMTDVAESSPPSAISFLPTEAG
NNDGRESGEGEKQLQKLVNHIQPHRAEALDLAIPLAVQKSPDMSKAHPLLGPPFPTHPSLALHHPSLP
RPPPLSHPSLLPPTSQSKDVPSSFHQSSSSSSSSSYPVPPQPPHPLPLPLLHYSMQQLFSRSLHHPQ
LPHLTPSRKDYLSAEPFFDFSSHPSTHPTFPPLPLLGPLDPSLARHSGKERERGMRGDGGMRGGLDGG
ELYLTAGGEGLSPCHLKKAKLMFFYARYPSSNTLKTYFPDVKFNRCVTSQMIKWFSNFREFFYIQMER
FARQAVREALTRDSQLRVGRDTELYRILNMHYNKSNVYQVPERFIEVSEVALREFYSAIWTGRDSDPC
WKKGIYKIICKLDSPVPDTFRLPGCPM

>Fugu*Prox2*
TTTGCAGTTATGGATTCCCCCTCTGACCTCTTTGAAGACTCCGGTCGCCAGATTCACACCTTTGCTTC
TACTCTCAGCTCCTCAGACCTCACAGGTGTTCAACTTCAGCCAAGCACACGTCATCCCCCTCCTGCTT
TCAGACCTCTGGGTTTCCCACTTATCCACAACCTCCTACATCCTGGTGGAGCAACCAGGAGAGTTGGA
GGGCAACTCAACCCAAACCTGAGAAGTCACAGGCACCTAGAGGAAGTCAACCTAGAGGAAAACATGCA
AGAGAGGGATGGACGTGTAGATCATGAGATGGATGGACTTATAGAAGGAGAGGACAGCTTGTTGGTGG
TTAAGAAGAGGCATAGTGACACTGTCCTCTTAGGTGAGTGGAGTCAGGACGTCCTGAAAGTGAAAAGG
GTAAGGAATGGAGAGCCAGAGGAGGGAGGCAAGAAGAACCAGGCAGGGAGGAAGCAGAAGAGACGAGA
GCGGGAGGAGCTGAAAGAGCAGCTGGAGGAGGCCAGAGAAAGGCTGCAGGCGCTGCAGGAGAAAGTAT
GGAAGGTTTTTGGAGAGAAGCACAGGGTGGAAGAGGAAAGGAGTCAACGTGGCAACAGAGATGCAGGA
GATGAGGGGATGGTGGAGGAAGAGGACATGTTTGAGGAAGATGATGGTGGAGACTTGGAAAAGGAAAA
CTTCTCCCTTCTTTCTGGTTCTCCTTTCGAGAACTTCCACAAACACAGAGAAGAGAAGCAGAAAGACA
CAGGGGGAAGAACGGAGAGGGAACGAGAGGGACAGTTGCACCTGGAGGGTGCGACGGACGGAGCAGAT
TTGACGATGGAAGCCGAGGACTGGAATGGGATGGAAGATGAAGGCGAGGAGGGAGGCCAGAAGTTTGC
TCAAGCGTTGAAGTTAGAGTTGGGCAGCGCCGTGGCACGGGTCATCGATCGAGTTCTCCGTCTTTACA
CCGAGATGACGGACGTTGCCGAGTCCTCCCCCCCTTCTGCCATCTCTTTCCTGCCAACTGAAGCTGGG
AACAATGACGGGAGGGAGAGCGGGGAAGGAGAGAAGCAGCTGCAAAAACTCGTCAACCACATCCAACC
ACATCGCGCTGAAGCATTGGACCTTGCGATACCACTGGCTGTTCAGAAGTCACCTGACATGTCAAAGG
CTCACCCCCTCCTTGGCCCGCCTTTCCCTACTCACCCGAGTCTCGCCCTGCACCACCCCTCTTTGCCT
CGCCCTCCTCCTCTTTCTCATCCGTCGCTCCTTCCACCCACCTCACAGTCCAAAGACGTCCCTTCCTC
CTTCCACCAGTCCTCCTCCTCCTCATCTTCCTCATCTTCCTACCCCGTGCCCCCCCAGCCTCCCCACC
CTCTCCCTCTCCCGCTCCTCCACTACTCCATGCAGCAGCTCTTCTCTCGTTCCCTTCACCACCCTCAG
CTGCCACACCTGACCCCGTCTCGCAAGGACTATTTGAGTGCCGAACCCTTTTTCGACTTCTCGTCTCA
CCCATCGACTCACCCAACTTTCCCGCCGCTCCCGTTGCTCGGGCCGCTGGACCCCTCCCTTGCACGCC
ACAGTGGAAAAGAGAGGGAGCGAGGAATGAGAGGAGATGGAGGAATGAGAGGAGGGCTAGACGGAGGA
GAGCTCTACCTGACAGCTGGGGGAGAAGGACTGTCTCCCTGTCACCTGAAGAAAGCCAAGCTCATGTT
CTTCTACGCCCGCTACCCAAGCTCCAACACTCTCAAGACGTACTTCCCCGACGTCAAGTTTAACCGCT
GTGTGACCTCCCAGATGATTAAATGGTTCAGCAACTTCAGAGAGTTCTTCTACATCCAGATGGAGCGA
TTTGCGCGACAAGCTGTCCGTGAGGCGCTCACTCGAGACAGCCAGCTGCGAGTGGGTCGTGACACAGA
GCTTTACCGCATTCTCAACATGCACTACAATAAAAGCAATGTCTACCAGGTCCCTGAAAGGTTCATCG
AAGTGTCAGAAGTGGCCCTGAGGGAGTTTTACTCAGCCATCTGGACTGGAAGAGACAGCGACCCCTGT
TGGAAGAAGGGGATATATAAGATTATTTGTAAGCTTGACAGTCCTGTACCAGACACCTTCAGACTTCC
GGGCTGCCCCATG

### 1.3.4 Fugu *prox3*

>FuguProx3
LPQHLSLRRPSARRLFSAQQVRSHLSAAVFALHPSEQFFSEMNLSHPDQNMHGAGDGCLEDDKPDVMF
PCFRRNVYDESLSSYSSGSIISQLLRKTIHSARALDESHFYLSSSGGADCSQEDQNSVSSKDSTMEAT
SPGAHVSTGVNPEQEHPVPDHLQAKRARVENIIRVMAGSPNSRQHGEIEKADTDAKDVRESREAYREN
KRKQRLPQHQEHGAGGPASRRPGSSNSDSCNAKDEECHKLKEQLHSMQKLLHQLQEKFLQVYNQEDPE
QNGRDEPEVNAQRDPMGENLASHYMSGEEDFDRKTSRASVECKGRMKAAGYMLQRESPNLQETLKQEL
SRAVNDCVDRVFKKATHTGLDVSPEQRMCSSPEVHISMVVDRKSQAAGCTQDQLQEEEGAEKKRSLEY
YESPSNQSPQDQTEALSLVVRKPEVTPLGAVAPTVKRPYPVHQTPFQFNYSTSLHDSQILEHLLKYGP
HSSFGGLPCIPPSMDRTSPDSVDLTWDAIAMRSKVTSGHLGHHGRPSALGAVTVDNLCLPHVKIECGE
LQSMAERNPYMSLNVNIQEGLTPSHLKKAKLMFFYTRYPSSNVLKTYFPDVKFNRCITSQLIKWFSNF
REFYYIQMEKFARQAIADGVSDVKDITVSRDSELFRALNMHYNKANDFHVPDRFLEVAEITLHEFYNA
IADTKDSDPSWKKAIYKVICKLDSDVPEEFKTSSY

>Fugu*prox3*
CTCCCGCAGCACCTCAGTCTTCGCAGGCCGTCAGCTCGGCGTCTCTTCTCAGCGCAACAGGTAAGGTC
CCATCTGTCTGCAGCCGTGTTTGCATTGCACCCGTCTGAACAGTTTTTCAGTGAGATGAACCTCAGCC
ATCCCGATCAGAACATGCACGGCGCCGGCGACGGCTGCCTGGAGGACGACAAACCCGATGTCATGTTC
CCCTGCTTTCGCAGAAACGTCTATGACGAGTCTCTCTCCTCCTACTCCAGCGGATCCATCATCTCTCA
GCTCCTGCGTAAGACGATCCACAGCGCGAGGGCGCTCGATGAAAGCCATTTCTACCTTTCCAGCTCTG
GCGGGGCAGATTGCAGCCAGGAGGACCAGAACAGTGTCTCCTCAAAGGACAGCACAATGGAGGCAACA
TCTCCTGGTGCTCATGTCTCCACTGGAGTCAACCCAGAGCAGGAGCATCCTGTGCCCGACCACCTGCA
GGCTAAGAGGGCCAGAGTGGAGAACATCATCAGGGTCATGGCAGGCTCTCCCAACAGCAGGCAACATG
GCGAGATTGAAAAGGCTGACACGGACGCCAAAGACGTAAGAGAGTCCAGAGAAGCCTACAGAGAGAAC
AAACGTAAGCAGAGGCTGCCGCAGCATCAGGAGCATGGTGCTGGAGGACCTGCAAGCAGGAGGCCCGG
AAGCAGCAACAGTGACTCCTGCAACGCCAAAGATGAGGAGTGTCACAAGCTGAAAGAGCAGCTGCACA
GCATGCAAAAGCTCCTGCATCAGCTCCAGGAGAAGTTCCTCCAGGTGTACAACCAGGAAGACCCCGAG
CAGAACGGCAGAGATGAGCCAGAGGTCAACGCTCAACGCGACCCCATGGGTGAAAACCTGGCATCGCA
CTACATGAGCGGCGAGGAGGATTTCGACAGGAAGACCAGCAGAGCGAGTGTGGAGTGTAAGGGCCGAA
TGAAAGCCGCTGGTTACATGCTCCAAAGAGAGAGTCCGAACCTGCAGGAGACTCTGAAGCAGGAGCTC
TCCAGGGCAGTGAACGACTGTGTGGACAGGGTGTTCAAGAAGGCCACCCACACGGGCCTAGATGTGTC
CCCTGAGCAGCGCATGTGCTCCTCACCTGAAGTGCACATCAGCATGGTGGTCGACAGGAAGAGCCAGG
CGGCCGGCTGCACCCAGGATCAACTGCAGGAAGAAGAGGGTGCGGAAAAGAAGCGTTCCCTGGAGTAC
TACGAAAGTCCCAGCAATCAGAGCCCGCAGGACCAGACGGAAGCGTTGTCACTGGTGGTCCGCAAACC
AGAGGTCACCCCTCTGGGCGCCGTCGCCCCAACAGTGAAGCGACCTTACCCTGTGCACCAAACACCCT
TCCAGTTCAATTACAGCACCTCTCTGCACGACAGCCAGATCCTAGAGCACCTTCTCAAGTACGGGCCA
CATTCCAGCTTCGGGGGACTCCCGTGCATCCCTCCGTCTATGGACAGGACCTCCCCGGACTCGGTGGA
CCTCACTTGGGACGCCATCGCCATGAGGTCCAAGGTGACGTCCGGCCACCTTGGCCACCACGGTCGTC
CCTCCGCCCTGGGGGCCGTGACAGTCGACAACCTGTGTCTTCCCCACGTGAAGATCGAGTGCGGTGAG
CTGCAGAGCATGGCCGAACGGAACCCTTACATGTCGCTCAATGTGAACATCCAGGAAGGTCTCACCCC
GAGCCATCTGAAGAAGGCCAAGCTAATGTTCTTCTACACCCGCTACCCCAGCTCCAACGTGCTGAAAA
CCTACTTCCCTGATGTGAAGTTCAACCGCTGCATCACCTCACAGCTGATCAAGTGGTTCAGCAACTTC
CGGGAGTTCTACTACATCCAGATGGAGAAGTTTGCCCGTCAGGCCATCGCTGACGGTGTCAGTGACGT
CAAAGACATTACAGTTAGCAGGGACTCAGAGCTGTTCCGGGCACTGAACATGCACTACAACAAAGCCA
ACGACTTCCACGTTCCAGACAGATTCCTGGAGGTGGCTGAGATCACCTTGCACGAGTTTTACAACGCC
ATCGCCGACACCAAGGATTCGGACCCCTCGTGGAAAAAAGCCATCTACAAGGTGATCTGTAAACTGGA
CAGCGACGTCCCGGAAGAATTCAAGACATCCTCTTAT

### 1.3.5 Fugu *prox4*

>FuguProx4
IDICISYGNRVLSTAFFFCSKTSVESMYSSSSYFHDGCTSDYLSSFQTDCLTSNLNFLDASVSLYEGS
EKQKCLPERDCCSPTDVISSSVCSSRPEFCPSRDAGPLLTKPDPSSGCGSSVGDWFLNSRREAKRARV
ENIIKGMSSTTGLSESMLGNKEVQALPEGCRSDRDSSVTSQRDACTDLHGEFDHSACATHDGWKRLLS
ATRAKSDRIELMTDVLKCELSRAIGRSIDSIFKSAPPLQTSPPDQGSLQAPVCSRGRQSRHAEDVQTE
ALSLVVPRAAQETTSDLSLQCGTRDAGPPLSESSLKKKAHQSPQGRPKVSSRSLRSLLVDPPPRPLIK
MESEATGNNYLYRLKVILTGERSATVQRKVVPPFSNVAEGLTTNHLKKAKLMFFYTRYPSSQVLKACF
HDVQLTRCVTSQLIKWFSNFREFYYIQMEKSARQAVAQGVTDVRSLAVERESGLFRALNTHYNKANDF
QVPQRFLEVAAITLREFYSAISTGEDRDPSWKKAIYK

>Fugu*prox4*
ATCGATATCTGCATATCTTATGGTAACCGTGTTCTGTCAACTGCTTTTTTCTTTTGTAGCAAAACTTC
CGTGGAAAGTATGTATTCCTCCAGCAGCTACTTCCATGACGGCTGCACCTCAGATTACCTTTCCTCTT
TTCAGACAGACTGTCTGACATCTAACTTAAACTTTCTGGACGCTTCTGTGAGTTTATATGAAGGCAGC
GAAAAGCAAAAATGTCTACCTGAGAGAGACTGCTGCAGCCCCACTGATGTTATTTCAAGCTCTGTCTG
TTCCAGCCGTCCAGAGTTCTGTCCCTCCAGAGACGCAGGACCGTTACTCACAAAACCAGATCCGTCTT
CTGGATGCGGTTCCTCCGTTGGCGACTGGTTTCTGAACAGTCGGCGTGAAGCCAAGCGCGCCAGAGTA
GAAAATATAATCAAAGGAATGAGCAGCACCACTGGTCTGTCAGAGAGCATGCTGGGAAATAAAGAGGT
CCAAGCCCTTCCTGAAGGCTGCAGATCTGATCGTGACTCCTCAGTGACATCCCAAAGAGATGCCTGCA
CAGATTTGCACGGTGAGTTTGACCACAGCGCGTGTGCGACGCACGACGGATGGAAGAGGCTTTTGAGC
GCCACCCGAGCCAAATCTGACAGGATTGAGCTGATGACAGACGTCCTGAAGTGTGAGCTTTCCAGAGC
CATCGGCAGGAGCATCGACTCCATATTCAAGAGCGCGCCGCCGCTGCAAACATCACCGCCCGACCAAG
GTTCCCTCCAGGCTCCGGTGTGCAGCAGGGGACGCCAGTCACGTCACGCTGAGGACGTTCAAACTGAA
GCGCTGTCGCTGGTGGTGCCGAGAGCTGCACAGGAAACCACTTCAGACCTCAGCCTTCAGTGTGGAAC
CAGAGATGCCGGACCTCCACTCTCTGAGTCCAGTCTGAAGAAGAAAGCCCATCAAAGCCCACAGGGGA
GACCCAAGGTGAGCTCCAGATCTCTCAGAAGCCTGCTGGTGGATCCGCCGCCTCGTCCTCTCATCAAG
ATGGAGTCTGAAGCCACAGGCAATAATTACCTCTACAGGCTGAAAGTTATCCTTACAGGAGAGCGATC
AGCTACGGTTCAAAGAAAAGTGGTCCCCCCGTTTAGTAACGTTGCC

# Appendix 2: Programs and Scripts

## 2.1 shiftcon

```
#! /usr/bin/tcsh -f
rm shi0* cshi0*
point7shift
foreach i (shi0*)
WlzCutObjToBox -i -o c$i -x 0,1391 -y -2,1042 $i
echo $i
end
TR_2DTo3D.csh cshi0 shifted.$1.wlz 0 100 1 Nrm
```

## 2.2 point7shift

```
#! /usr/bin/perl -w

$double0 = "00";
$single0 = "0";
$image_number = 7;

while ($image_number<400) {

   if ($image_number < 100) {$zero_number = "$single0$image_number";}
   if ($image_number < 10) {$zero_number = "$double0$image_number";}
   if ($image_number > 99) {$zero_number = "$image_number";}
         `WlzAffineTransformObj -y 1 -L -o shi0$zero_number.wlz pix0$zero_number.wlz`;
         $image_number += 8;
         print "shift0$zero_number.wlz\n";
}
$image_number = 0;
while ($image_number<400) {

   if ($image_number < 100) {$zero_number = "$single0$image_number";}
   if ($image_number < 10) {$zero_number = "$double0$image_number";}
   if ($image_number > 99) {$zero_number = "$image_number";}
         `WlzAffineTransformObj -y 0.7 -L -o shi0$zero_number.wlz pix0$zero_number.wlz`;
         $image_number += 8;
         print "shift0$zero_number.wlz\n";
}
$image_number = 1;
while ($image_number<400) {

   if ($image_number < 100) {$zero_number = "$single0$image_number";}
   if ($image_number < 10) {$zero_number = "$double0$image_number";}
   if ($image_number > 99) {$zero_number = "$image_number";}
         `cp pix0$zero_number.wlz shi0$zero_number.wlz`;
         $image_number += 8;
```

```
        print "shift0$zero_number.wlz\n";
}

$image_number = 2;
while ($image_number < 400) {
   if ($image_number < 100) {$zero_number = "$single0$image_number";}
   if ($image_number < 10) {$zero_number = "$double0$image_number";}
   if ($image_number > 99) {$zero_number = "$image_number";}
        `WlzAffineTransformObj -y -0.7 -L -o shi0$zero_number.wlz pix0$zero_number.wlz`;
        $image_number += 8;
        print "shift0$zero_number.wlz\n";
}

$image_number = 3;
while ($image_number<400) {

   if ($image_number < 100) {
$zero_number = "$single0$image_number";
}
   if ($image_number < 10) {$zero_number = "$double0$image_number";}
   if ($image_number > 99) {$zero_number = "$image_number";}
        `WlzAffineTransformObj -y -1 -L -o shi0$zero_number.wlz pix0$zero_number.wlz`;
        $image_number += 8;
        print "shift0$zero_number.wlz\n";
}
$image_number = 4;
while ($image_number<400) {

   if ($image_number < 100) {$zero_number = "$single0$image_number";}
   if ($image_number < 10) {$zero_number = "$double0$image_number";}
   if ($image_number > 99) {$zero_number = "$image_number";}

        `WlzAffineTransformObj -y -0.7 -L -o shi0$zero_number.wlz pix0$zero_number.wlz`;
        $image_number += 8;
        print "shift0$zero_number.wlz\n";
}
$image_number = 5;
while ($image_number<400) {

   if ($image_number < 100) {$zero_number = "$single0$image_number";}
   if ($image_number < 10) {$zero_number = "$double0$image_number";}
   if ($image_number > 99) {$zero_number = "$image_number";}
        `cp pix0$zero_number.wlz shi0$zero_number.wlz`;
        $image_number += 8;
        print "shift0$zero_number.wlz\n";
}

$image_number = 6;
while ($image_number<400) {

   if ($image_number < 100) {$zero_number = "$single0$image_number";}
   if ($image_number < 10) {$zero_number = "$double0$image_number";}
   if ($image_number > 99) {$zero_number = "$image_number";}
        `WlzAffineTransformObj -y 0.7 -L -o shi0$zero_number.wlz pix0$zero_number.wlz`;
        $image_number += 8;
        print "shift0$zero_number.wlz\n";
}
exit;
```

## 2.3 singleshift

The files referred to outr1.wlz to outtr8.wlz are matrix files consisting of the coordinates calculated for the single transformation.

```
#! /usr/bin/perl -w

$double0 = "00";
$single0 = "0";
$image_number = 7;

while ($image_number<400) {

   if ($image_number < 100) {$zero_number = "$single0$image_number";}
   if ($image_number < 10) {$zero_number = "$double0$image_number";}
   if ($image_number > 99) {$zero_number = "$image_number";}
         `WlzAffineTransformObj -t outtr1.wlz -L -o shi0$zero_number.wlz pix0$zero_number.wlz`;
         $image_number += 8;
         print "shift0$zero_number.wlz\n";
}
$image_number = 0;
while ($image_number<400) {

   if ($image_number < 100) {$zero_number = "$single0$image_number";}
   if ($image_number < 10) {$zero_number = "$double0$image_number";}
   if ($image_number > 99) {$zero_number = "$image_number";}
         `WlzAffineTransformObj -t outtr2.wlz -L -o shi0$zero_number.wlz pix0$zero_number.wlz`;
         $image_number += 8;
         print "shift0$zero_number.wlz\n";
}
$image_number = 1;
while ($image_number<400) {

   if ($image_number < 100) {$zero_number = "$single0$image_number";}
   if ($image_number < 10) {$zero_number = "$double0$image_number";}
   if ($image_number > 99) {$zero_number = "$image_number";}
         `WlzAffineTransformObj -t outtr3.wlz -L -o  shi0$zero_number.wlz
pix0$zero_number.wlz`;
         $image_number += 8;
         print "shift0$zero_number.wlz\n";
}

$image_number = 2;
while ($image_number < 400) {
   if ($image_number < 100) {$zero_number = "$single0$image_number";}
   if ($image_number < 10) {$zero_number = "$double0$image_number";}
   if ($image_number > 99) {$zero_number = "$image_number";}
         `WlzAffineTransformObj -t outtr4.wlz -L -o shi0$zero_number.wlz pix0$zero_number.wlz`;
         $image_number += 8;
         print "shift0$zero_number.wlz\n";
}

$image_number = 3;
while ($image_number<400) {
```

```
   if ($image_number < 100) {
$zero_number = "$single0$image_number";
}
   if ($image_number < 10) {$zero_number = "$double0$image_number";}
   if ($image_number > 99) {$zero_number = "$image_number";}
         `WlzAffineTransformObj -t outtr5.wlz -L -o shi0$zero_number.wlz pix0$zero_number.wlz`;
         $image_number += 8;
         print "shift0$zero_number.wlz\n";
}
$image_number = 4;
while ($image_number<400) {

   if ($image_number < 100) {$zero_number = "$single0$image_number";}
   if ($image_number < 10) {$zero_number = "$double0$image_number";}
   if ($image_number > 99) {$zero_number = "$image_number";}

         `WlzAffineTransformObj -t outtr6.wlz -L -o shi0$zero_number.wlz pix0$zero_number.wlz`;
         $image_number += 8;
         print "shift0$zero_number.wlz\n";
}
$image_number = 5;
while ($image_number<400) {

   if ($image_number < 100) {$zero_number = "$single0$image_number";}
   if ($image_number < 10) {$zero_number = "$double0$image_number";}
   if ($image_number > 99) {$zero_number = "$image_number";}
         `WlzAffineTransformObj -t outtr7.wlz -o shi0$zero_number.wlz pix0$zero_number.wlz`;
         $image_number += 8;
         print "shift0$zero_number.wlz\n";
}

$image_number = 6;
while ($image_number<400) {

   if ($image_number < 100) {$zero_number = "$single0$image_number";}
   if ($image_number < 10) {$zero_number = "$double0$image_number";}
   if ($image_number > 99) {$zero_number = "$image_number";}
         `WlzAffineTransformObj -t outtr8.wlz  -L -o shi0$zero_number.wlz
pix0$zero_number.wlz`;
         $image_number += 8;
         print "shift0$zero_number.wlz\n";
}
exit;
```

## 2.4 remrecon

```
#! /usr/bin/perl
#recon options and number of machines
#divide sections equally among machines
#rsh recon on all machines
#output names as box followed by the range of sections

use Getopt::Std;
$gain = 1000;
$adjust = 15;
$dir = `pwd`;
$power=" ";
#getopt first and last plane, depth of boxes,
getopts ('Fhf:l:d:hr:g:pa:');

if (!getopts('Fhf:l:d:hr:g:pa:')) {
      print "Usage remrecon -Ffldgmhp recon -h for help\n";
      exit;
}

if ($opt_F) {
      $type = "-F";
}

if ($opt_f) {
      $first = $opt_f;
}

if ($opt_l) {
      $last = $opt_l;
}

if ($opt_d) {
      $depth =$opt_d;
}

if ($opt_g) {
      $gain =$opt_g;
}

if ($opt_m) {
      $adjust =$opt_m;
}

if ($opt_h) {
      print" Usage: remrecon -[Ffldgm]\n";
      print" -F flouresence data and p3\n";
      print" -f first section\n";
      print" -l last section\n";
      print" -d depth of boxes (0 means a single plane)\n";
      print" -g gain (default 1000)\n";
      print" -m range of adjust values (default=15)\n";
      print" -h this help message\n";
      exit;
```

```
}

if ($opt_r) {
        $remotes = $opt_r;
        }

if ($opt_p) {
  $power = "-p $opt_p";
}

if ($opt_a) {
  $adjust = $opt_a;
}

        $boxheight = ($last - $first)/($remotes+1);
        $height =$boxheight/$depth;
        $boxheight = (int $height) * $depth;
        if ($boxheight<$depth) {
                $boxheight = $depth;
                }

        print $boxheight;
        $start = $first;
        $end = $first + $boxheight;

        while ($remotenames<$remotes) {
                $remotenames ++;
                print "What is the name of remote number $remotenames?\n";
                $name = <STDIN>;
                chomp $name;
                system "rsh $name 'cd $dir ;mrecon $type $power -f$start -l$end -d$depth -g$gain -
a $adjust'&";
                $start ++;
                $start += $boxheight;
                $end ++;
                $end += $boxheight;
                if ($end>$last) {
                   $end = $last;
                }
                }

system "mrecon $type $power -f$start -l$last -d$depth -g$gain -a $adjust";
exit;
```

## 2.5 mrecon

```
#! /usr/bin/perl

# reconstruct number of 12 bit sections of x planes
# measure grey range across all boxes
# setgreyrange and convert to 8 bit
# join all the boxes
# any errors/suggestions email robertr

use Getopt::Std;
$gain =1000;
$adjust=15;
#getopt first and last plane, depth of boxes,

getopts('Ff:l:d:h:g:m:p:a:');

if (!getopts('Ff:l:d:h:g:m:p:a:')) {
        print "Usage recon -Ffldgmhp recon -h for help\n";
        exit;
}

if ($opt_F) {
        $type = "-F";
}

if ($opt_f) {
        $first = $opt_f;
} else {
        print "must specify first section\n";
        print "recon -h for help\n";
        exit;
}

if ($opt_l) {
        $last = $opt_l;
} else {
        print "must specify last section\n";
        print "recon -h for help\n";
        exit;
}

if ($opt_d) {
        $depth =$opt_d;
} else {
        print "must specify depth of box\n";
        print "recon -h for help\n";
        exit;
}

if ($opt_g) {
        $gain =$opt_g;
}

if ($opt_m) {
```

```
        $adjust =$opt_m;
}
if ($opt_h) {
        print" Usage: recon -[Ffldgmp]\n";
        print" -F flouresence data\n";
        print" -p power ";
        print" -f first section\n";
        print" -l last section\n";
        print" -d depth of boxes (0 means a single plane)\n";
        print" -g gain (default 1000)\n";
        print" -m range of adjust values (default=15)\n";
        print" -h this help message\n";
        print" -a adjsut value\n";
        exit;
}
if ($opt_p) {
    $power="-p$opt_p";
}

if ($opt_a) {
 $adjust = $opt_a;
}

$topbox = $first;
$basebox = $first + $depth;

if ($basebox >$last) {
        $basebox = $last;
        }
$box = 1;
$end = 0;

while ($end <1) {

        if ($basebox >= $last) {
        $basebox = $last;
        $end = 1;
        }
        print "$topbox\n";
        system "OptRecon -v $type $power -f $topbox -l $basebox -i aln -o box$topbox-
$basebox.wlz -g $gain -m $adjust -a $adjust";
        $topbox = $topbox + $depth + 1;
        $basebox = $basebox + $depth + 1;
        $box ++;

}

exit;
```

## 2.6 zebedee

Zebedee was written by Robert Bryson-Richardson and Colin Semple. Certain parts of the code, as indicated, are taken from scripts by Martin Taylor.

```
#!/net/ledaig/export/usr5/wrkgrp/bioinf/bin/perl -w
#script to find zebrafish homologue coding seqs in shotgun read data
#takes a single fasta formatted protein sequence and a %id cutoff and length of match as input

use Getopt::Std;

########parse args#########
# set defaults

getopt("silpd", \%args);
our $in = $args{s} || usage(); #ie execute usage to get value of in
our $id_cutoff = $args{i} || 50;
our $l_cutoff = $args{l} || 25;
our $in_type = $args{p} || T;
our $datab = $args{d} || Dr_wgs;
our $org = $datab;
#our $far = $in.".reads";

if (!-e "/net/glenfarclas/export/data1/zebedee/Dr_wgs/$datab.nal") {
        die "database $datab does not exist\n";
}

# if assembly has already been done -only blast against updates
# set $datab to more recent of two update s in Dr_wgs if Dr or Mm

if (-e $in.".$datab") {
        @files = `cd  /net/glenfarclas/export/data1/zebedee/Dr_wgs; ls -t $datab.update*.nsq`;
        if (defined $files[0]) {
        $files[0] =~ s/\.nsq//;
        $datab = "$files[0]";
        chomp $datab ;
        } else {
        print "There are no updates for $datab\n";
        exit;
        }
}

print "in = $in database = $datab id_cutoff = $id_cutoff l_cutoff = $l_cutoff in_type = $in_type\n";

open(IN,"$in") || die "Cannot open $in: $!\n";
open(OUT1,">$in\.$org") || die "Cannot open OUT\n";
my $tm = localtime;
print OUT1 "zebedee started at $tm with:\n";
print OUT1 "percentage identity cutoff = $id_cutoff, BLAST hit alignment length cutoff = $l_cutoff,
input file type = $in_type, and input sequence =\n";
my $seq = "";
while (<IN>) {
        $seq .= $_;
```

```
}
print OUT1 "$seq\n";

########make CONSED dir tree#########

$dir = $in."\.".$org."_dir";
$edir = $dir."/edit_dir";
$pdir = $dir."/phd_dir";
$cdir = $dir."/chromat_dir";
$vdir = $dir."/vector_dir";
mkdir $dir, 0755 or die "can't mkdir $dir: $!\n";
mkdir $edir, 0755 or die "can't mkdir $dir: $!\n";
mkdir $pdir, 0755 or die "can't mkdir $dir: $!\n";
mkdir $cdir, 0755 or die "can't mkdir $dir: $!\n";
mkdir $vdir, 0755 or die "can't mkdir $dir: $!\n";

link $in,$edir."/".$in;
chdir "$edir" or die "Can't cd to $edir: $!\n";

########find matching read hsps using BLAST#########

if ($in_type eq "T") {
        `/opt/bioinf/blast2.1.3/blastall -p tblastn -d
/net/glenfarclas/export/data1/zebedee/Dr_wgs/$datab -F F -i $in > $in\.blastout`;
}
else {
        `/opt/bioinf/blast2.1.3/blastall -p blastn -d
/net/glenfarclas/export/data1/zebedee/Dr_wgs/$datab -F F -i $in > $in\.blastout`;
}
my @hsrs = split /\n/,`/opt/bioinf/scripts/hsr_parse.pl $in\.dr $id_cutoff $l_cutoff`;
foreach my $item (@hsrs) {
        $item =~ s/(.*?\)\s+)(\S+)(.*)/$2/; #reduce hsps to read names
}

########remove duplicate reads#########

my %seen = ();
my @todo = ();
foreach $item (@hsrs) {
  unless ($seen{$item}) { #ie if we have not seen it before
    $seen{$item} = 1;
    push(@todo, $item);
  }
}
print "todo = @todo\n";

########retrieve todo list using FASTACMD#########

$fa_r = $in.".reads";
open(OUT2,">$fa_r") || die "Cannot open OUT2\n";
foreach $t (@todo) {
        my $rseq = `/opt/bioinf/blast2.1.3/fastacmd -d
/net/glenfarclas/export/data1/zebedee/Dr_wgs/$datab -s $t`;
        $rseq =~ s/lcl\|//; #get rid of annoying header prefix
        print OUT2 "$rseq";
        if ($org eq "Dr_wgs") {
                get_trace ($t);
                my $temp1 = $t.".scf";
```

```
                my $temp2 = "../chromat_dir/".$temp1;
                `mv $temp1 $temp2`; #move trace file from edir to cdir
                    }

}
close OUT2;

#split .reads into single fasta files if not Dr_wgs

if ($org ne "Dr_wgs") {
        stfa ($fa_r);
        foreach $t (@todo) {

                my $temp1 = "$t";
                my $temp2 = "$t.trace";
                `/opt/local/genome/bin/mktrace $temp1 $temp2; rm $temp1 $temp2`;
                `mv $temp2.phd.1  ../phd_dir/$t.phd.1`;
                }
}

my $readno = @todo;
print OUT1 "Set of $readno new reads generated from $in: @todo\n\n";

########assemble reads using PHRAP#########

$ENV{CONSED_HOME}="/opt/bioinf/consed11/"; #set envir var for consed
`/opt/bioinf/consed11/bin/phredPhrap $fa_r`;
my $fa_c = $fa_r.".contigs"; # phrap ctgs
my $fa_s = $fa_r.".fasta.screen.singlets"; # phrap singlets
my $c_no = `grep -c '>' $fa_c`; #nos ctgs
chomp ($c_no);
my $s_no = 0; #nos singlets
open(TEMP1,"$fa_s") || die "Cannot open $fa_s\n";
open(TEMP2,">ctgs_+_sglts") || die "Cannot open ctgs_+_sglts\n";
while (<TEMP1>) {
        if ($_ =~ /^>/) {
                $_ =~ s/(^>\S+)(\.scf.*)/$1/; #reduce singlet names
                $s_no++;
        }

#get name from singlet read
        if ($_ =~/\.trace/) {
                $_ =~ s/\.trace.+//;
        }

        print TEMP2 "$_";
}
close TEMP1;
close TEMP2;
print OUT1 "Set of $c_no contigs and $s_no singlets generated from both new and old reads by phrap
...\n";
`cat $fa_c >> ctgs_+_sglts`;

########order ctgs and sglts using BLAST#########

$ENV{NCBI}="/opt/bioinf/blast2.1.3/"; #set envir var for formatdb
`/opt/bioinf/blast2.1.3/formatdb -i ctgs_+_sglts -p F`;
if ($in_type eq "T") {
```

```
        `/opt/bioinf/blast2.1.3/blastall -p tblastn -d ctgs_+_sglts -i $in -F F> $in\.ctgs_+_sglts`;
}
else {
        `/opt/bioinf/blast2.1.3/blastall -p blastn -d ctgs_+_sglts -i $in -F F> $in\.ctgs_+_sglts`;
}
my $new_id_cut = $id_cutoff - 10; # more liberal matching
@hsrs = split /\n/,`/opt/bioinf/scripts/hsr_parse.pl $in\.ctgs_+_sglts $new_id_cut $l_cutoff`;
my @sbjcts = ();
my @qhsp = (); #hsp coords
my @shsp = (); #hsp coords
my @ids = ();
my @fr = ();
my %tosort;
$x = 0;
foreach $item (@hsrs) { #get all hsrs
        $item =~ s/(.*?\()//; #get rid of queryname
        $item =~ s/[\(\)]//g; #get rid of brackets
        ($qhsp[$x], $sbjcts[$x], $shsp[$x], $fr[$x], $ids[$x], $blah, $blah, $blah) = split /\s+/,$item;
        ($temp1, $blah) = split /-/,$qhsp[$x]; #get start of hsp in query
        $temp2 = $shsp[$x]."bp of ".$sbjcts[$x]." matches ".$qhsp[$x]." of ".$in." (frame =
".$fr[$x].", %identity = ".$ids[$x].")"; #hsr desc
        $tosort{$temp2} = $temp1; #assign hsp start coord as value
        $x++;
}

print OUT1 "\nBased on matches to $in the suggested order of phrap contigs and singlets is:\n\n";

#loop to sort hash numerically by values and print keys:
foreach $item (sort { $tosort{$a} <=> $tosort{$b} } keys %tosort) {
        print OUT1 "$item\n";
}

$tm = localtime;
print OUT1 "\nZebedee terminated at $tm\n";
print OUT1 "\nNB high identity matches to the query sequence do not demonstrate that the shotgun
sequence encodes the same gene\n";
print OUT1 "Contigs and singlets matching the query sequence should be compared to a protein
database\n";
close IN;
close OUT1;


sub get_trace {
# sub get trace taken from a script written by Martin Taylor
#dumps trace file for a zebrafish read from trace.ensembl.org
#NB server needs to be prodded (ie first wget call) for .scf file to be created

my $in = shift;
`/opt/local/bin/wget \"http://trace.ensembl.org/perl/traceview?tracedb=1&traceid=$in\"`;
`rm *traceview*`;
$in = $in.".scf.gz";
`/opt/local/bin/wget \"http://trace.ensembl.org/tmp/$in\"`;
print "wgot $in\n";
`gunzip $in`;
}

sub usage {
```

```
   print "Zebedee: a BLAST/phredPhrap wrapper\nColin.Semple\@hgu.mrc.ac.uk + Robert.Bryson-
Richardson\@hgu.mrc.ac.uk 2001\nusage: zebedee.pl -s <seqfile> [flags]\n";
   print "-s a mandatory FASTA formatted sequence file\n-d database to search Dr_wgs, Fr_asm,
Fr_wgs (default = Dr_wgs)\n";
   print "-i percentage identity threshold for matching shotgun reads (default = 50)\n-l alignment
length threshold for matching shotgun reads (default = 25)\n";
   print"-p sequence type: T for protein and F for nucleotide (default = T)\n";
exit;
}

sub stfa {
#splits multiple fasta a file into single fasta files - from stfa written by Martin Taylor
my $file = shift;
open (BLAST,"$file");

foreach $lin (<BLAST>){
      if ($lin =~ /^>/){
            chomp $lin;
            $go++;
            @ti = split /\s+/, $lin;
            $tii = $ti[0];
            $tii =~ s/^>//;
            open (FO, ">$tii") || die "open failed $tii\n";
            print FO ">$tii\n";
            next;
      }
      if ($go >= 1){
            if ($lin =~ /^\w+/){
                  print FO "$lin";
            }
            else { $go = 0; close FO;}
      }
}
}
```